From Knowledge Generation To Science-based Innovation

INESCTEC
TECHNOLOGY & SCIENCE
ASSOCIATE LABORATORY

30 1985 THIRTY YEARS 2015

FROM KNOWLEDGE GENERATION TO SCIENCE-BASED INNOVATION

Raia

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 Tecnlóxico 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

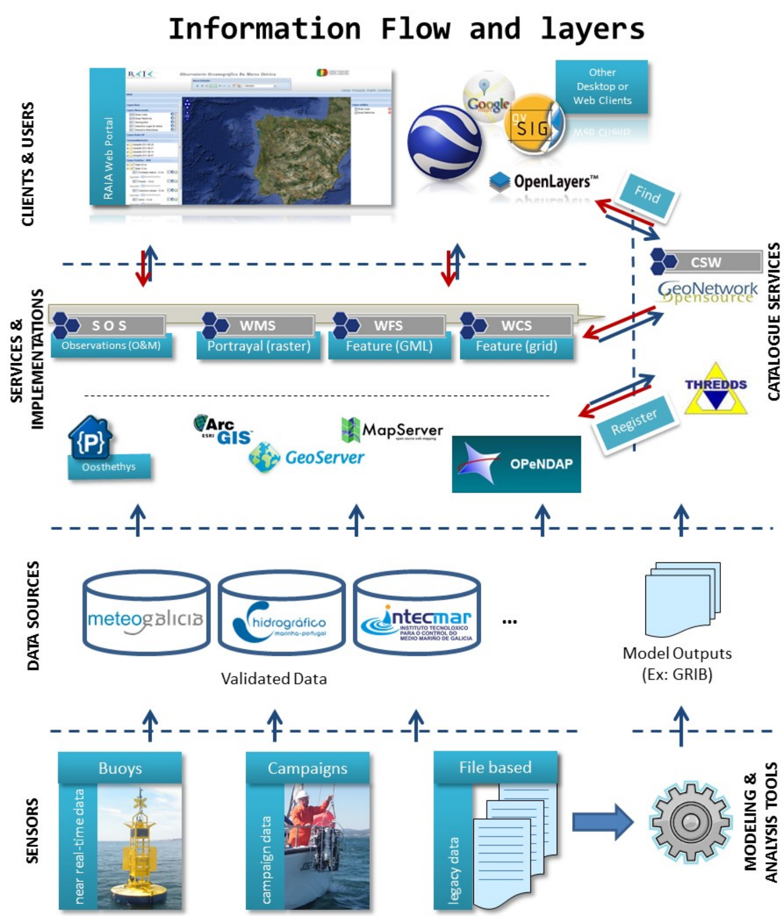
Research and Technological Development | Technology Transfer and Valorisation | Advanced Training | Consulting
Pre-incubation of Technology-based Companies

COOPERACIÓN TRANSFRONTERIZA ESPAÑA – PORTUGAL COOPERAÇÃO TRANSFRONTEIRIÇA

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



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.

FROM KNOWLEDGE GENERATION TO SCIENCE-BASED INNOVATION

1985 THIRTY YEARS 2015

INESCTEC TECHNOLOGY & SCIENCE | ASSOCIATE LABORATORY

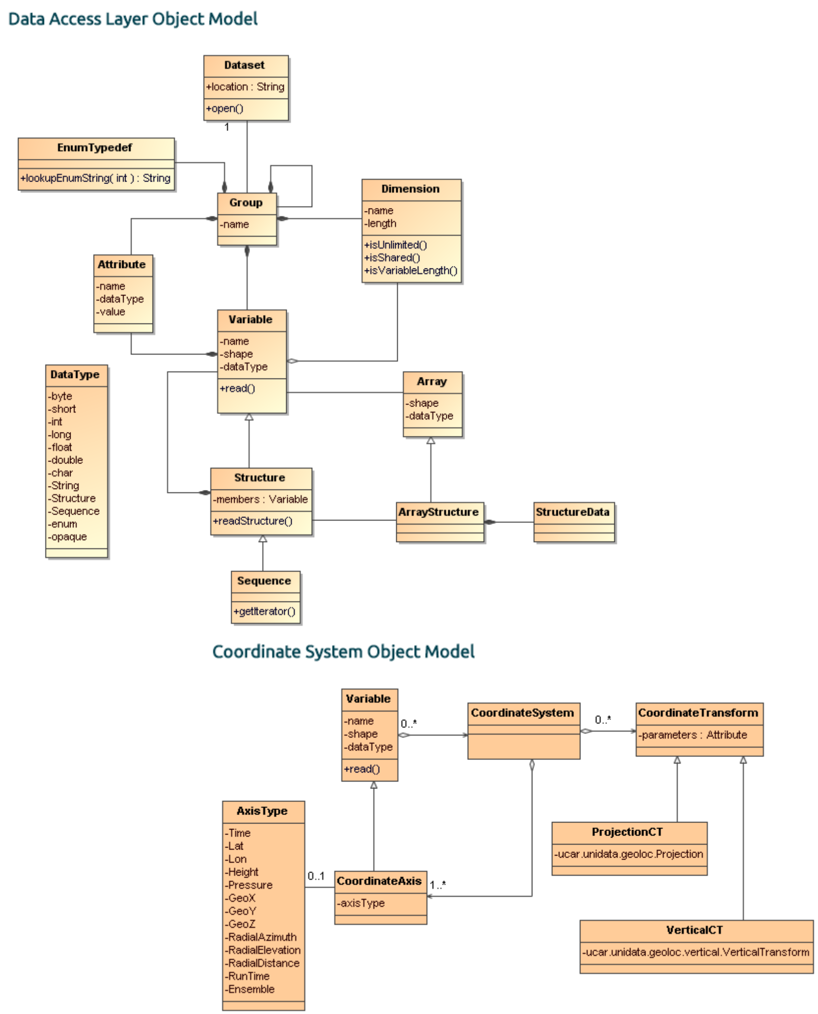
Methods and Implementation

- 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:

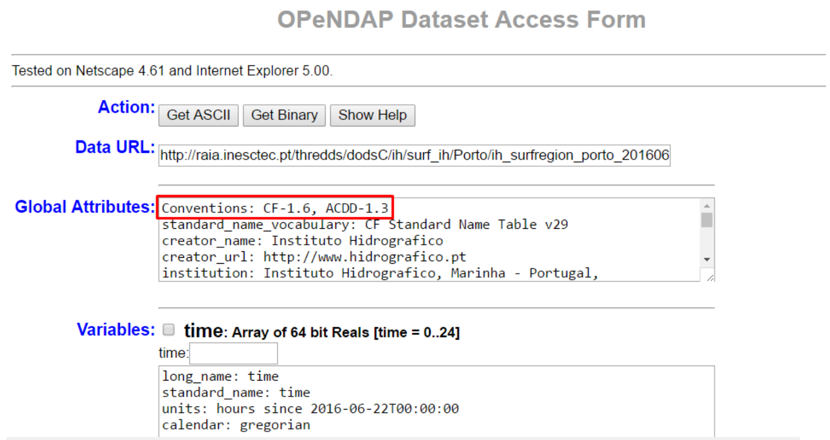
http://cfconventions.org/standard-names.html

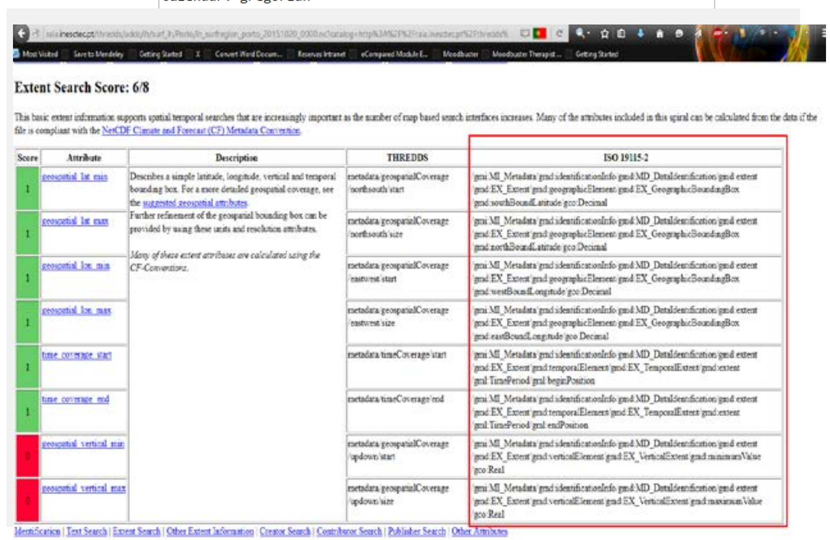


(1) [surface] [component] standard_name [at surface] [in medium] [due to process] [assuming condition]

Implementation Example 1: THREDDS Data Servers

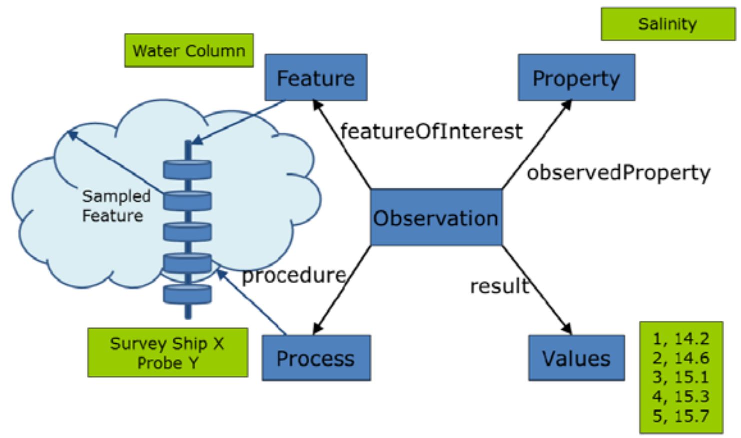
- Partners agreed to use at least CF on version 1.4 (some using CF 1.6)
- THREDDS Data Server implements several standard APIs to access data (WMS, WFS, WCS, etc)
- Metadata accessible according to ISO19115-2
- Simple interface to check completeness of the metadata





Implementation Example 2: SOS

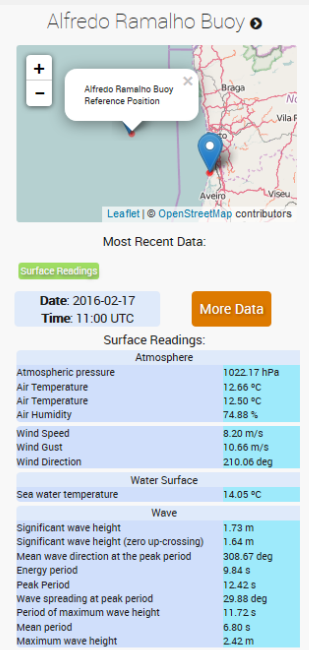
- Standard to provide access to time series from sensors and sensor systems;
- Goal is to be able to query data from sensors independently of which sensor is providing it;
- Facilitate data tailoring and fusion for (real-time) observations, enabling cross-domain analysis.

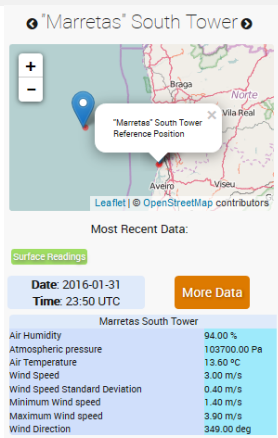


- Advantages:
  - Richer API
  - Includes reference to Feature of Interest and Process (Sensor, Lab Process, …)
- Disadvantage:
  - Not so efficient for bulkier data (ex: forecasts)

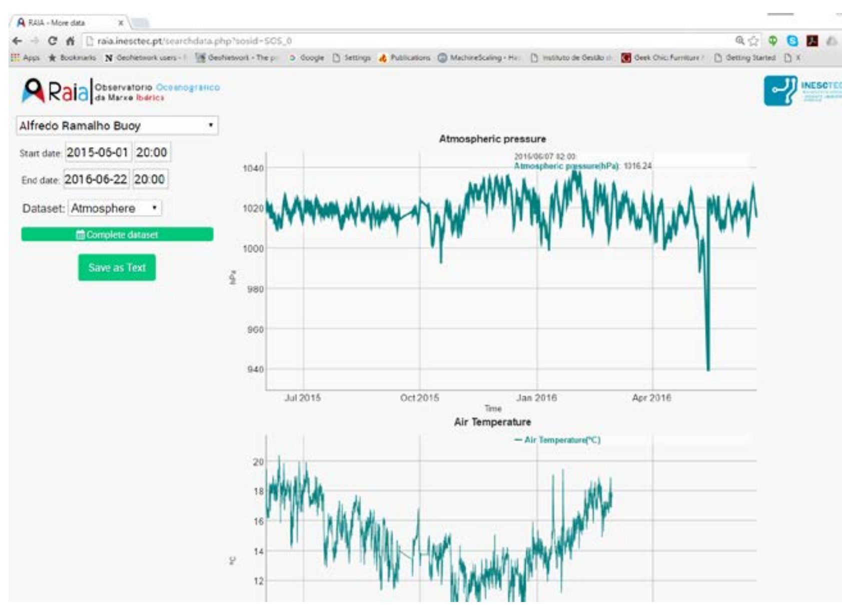
Implementation Example 2: SOS

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

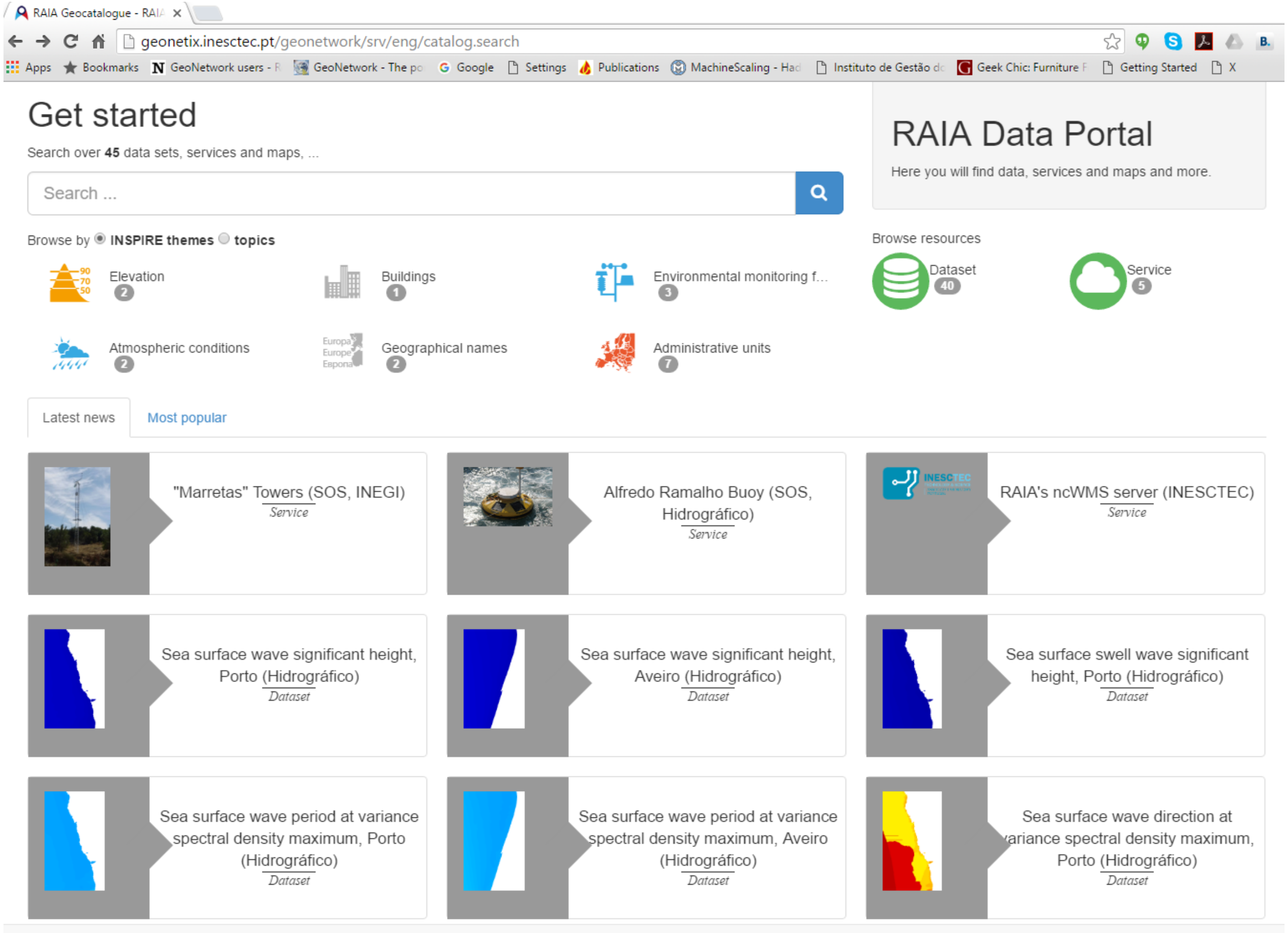




- Example for a more thorough exploration of a sensor time series:



Example of automatic harvesting of service metadata

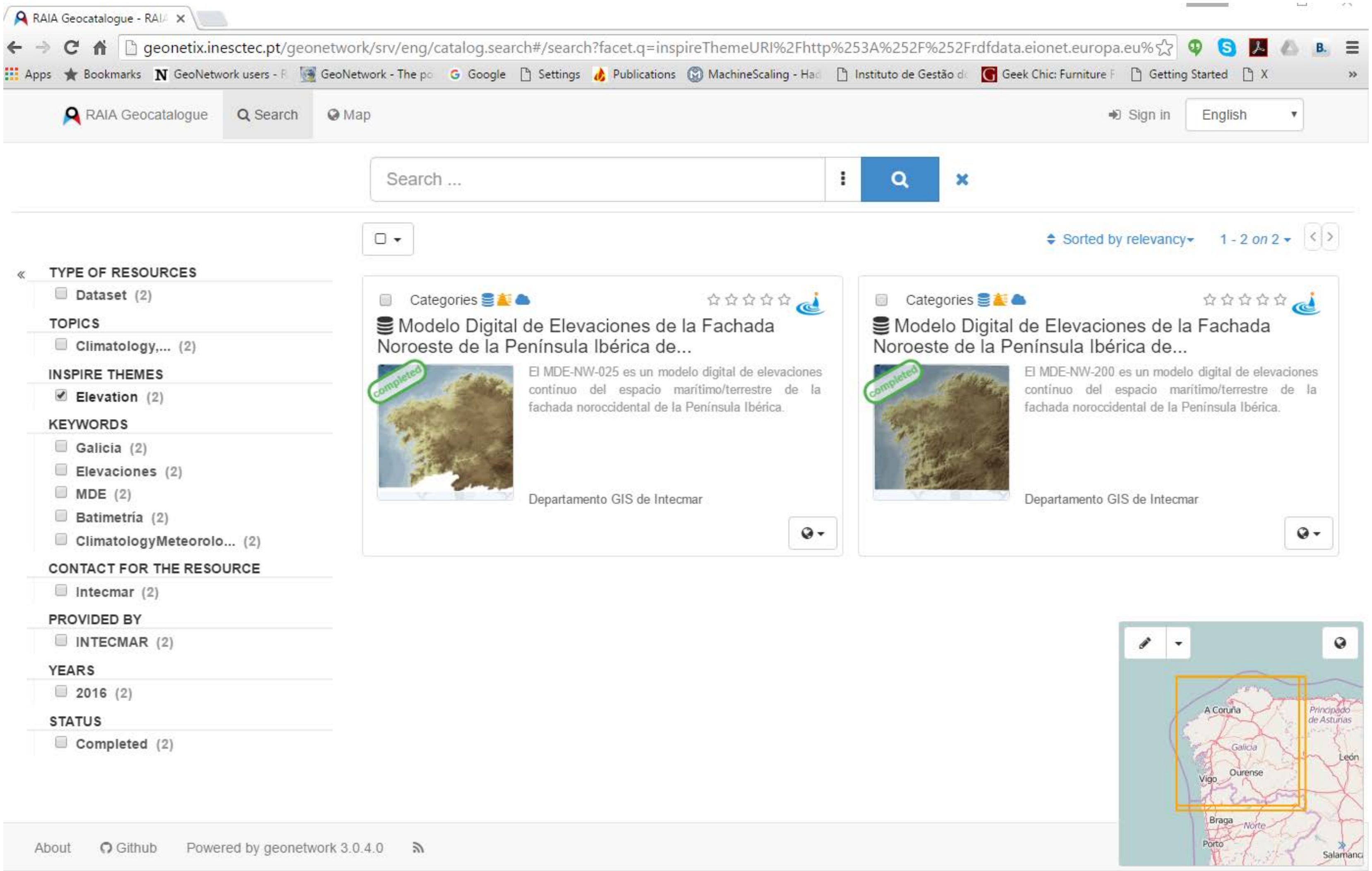


FROM KNOWLEDGE GENERATION TO SCIENCE-BASED INNOVATION

1985 THIRTY YEARS 2015

INESCTEC TECHNOLOGY & SCIENCE | ASSOCIATE LABORATORY

Example of automatic harvesting of dataset metadata



FROM KNOWLEDGE GENERATION TO SCIENCE-BASED INNOVATION

1985 THIRTY YEARS 2015

INESCTEC TECHNOLOGY & SCIENCE ASSOCIATE LABORATORY

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.

FROM KNOWLEDGE GENERATION TO SCIENCE-BASED INNOVATION

1985 THIRTY YEARS 2015

INESCTEC TECHNOLOGY & SCIENCE | ASSOCIATE LABORATORY