

# Semantic-Web and Oceanography: ISPRA Marine Observation through Linked Open Data

Marco Picone, Elio Giulianelli, Gabriele Nardone and Carlo Cipolloni

Italian Institute for Environmental Protection and Research, ISPRA, Italy



dct:description

The Linked Open Data Portal in ISPRA

<http://dati.isprambiente.it>

## CHALLENGES

*From strings to things*

The application of semantic technologies to environmental and oceanographic knowledge in order to guarantee a suitable data access, avoiding chaotic information, is nowadays not an utopia and in the last decade a great number of agreements, guidelines, laws, project have been developed. In this framework, Linked Open Data (LOD) philosophy is based on the idea of using specific web resources and connecting them through similarity or equivalence mechanisms, facilitating the improvement of the initial information set, according to rules expressed by Tim Berners-Lee in 2006.



## INPUT

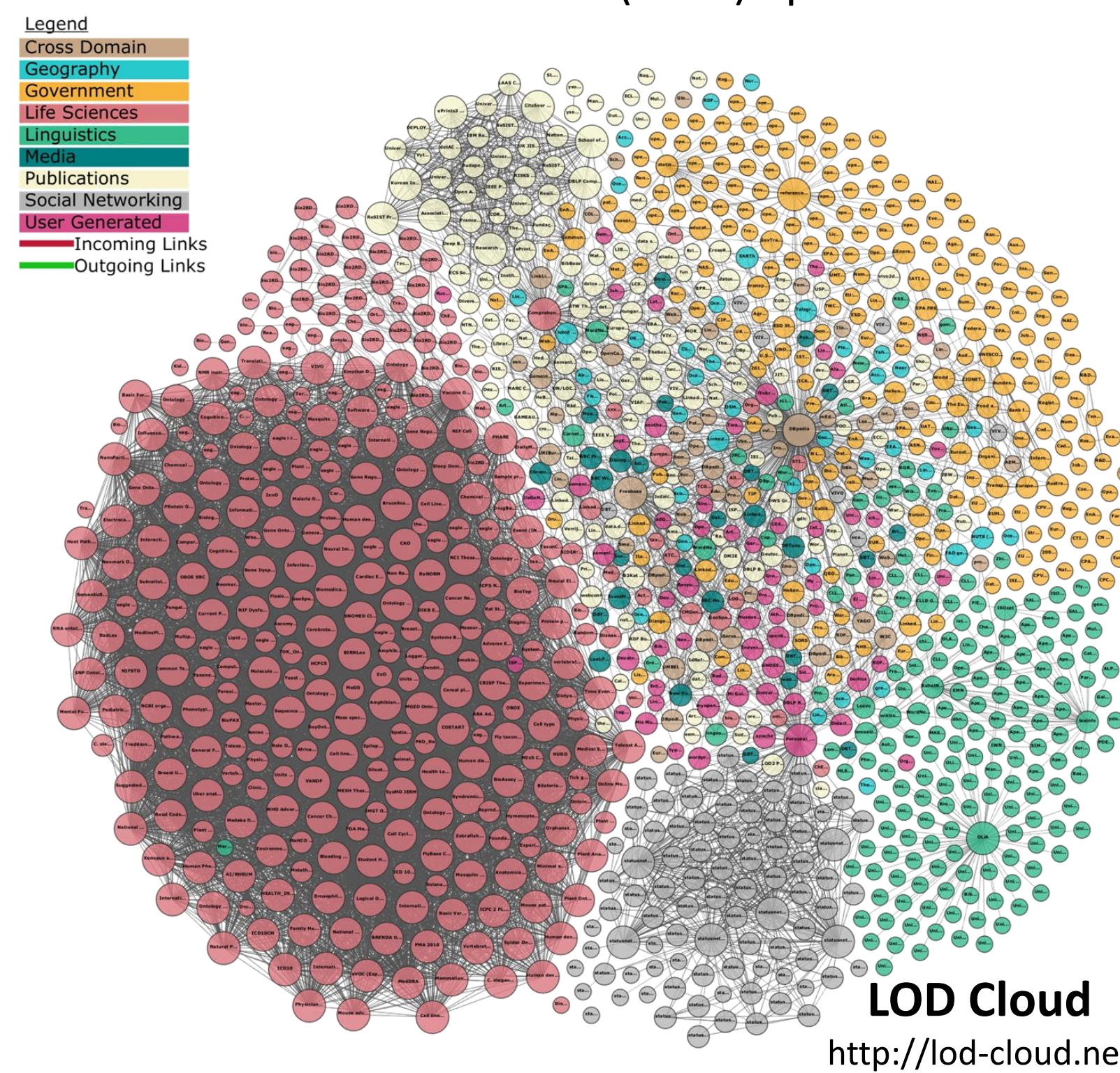
GML  
XML  
CSV  
Plain text  
XLS  
...



## THE PROJECT

Visit us on <http://dati.isprambiente.it>  
or look for us on Linked data cloud

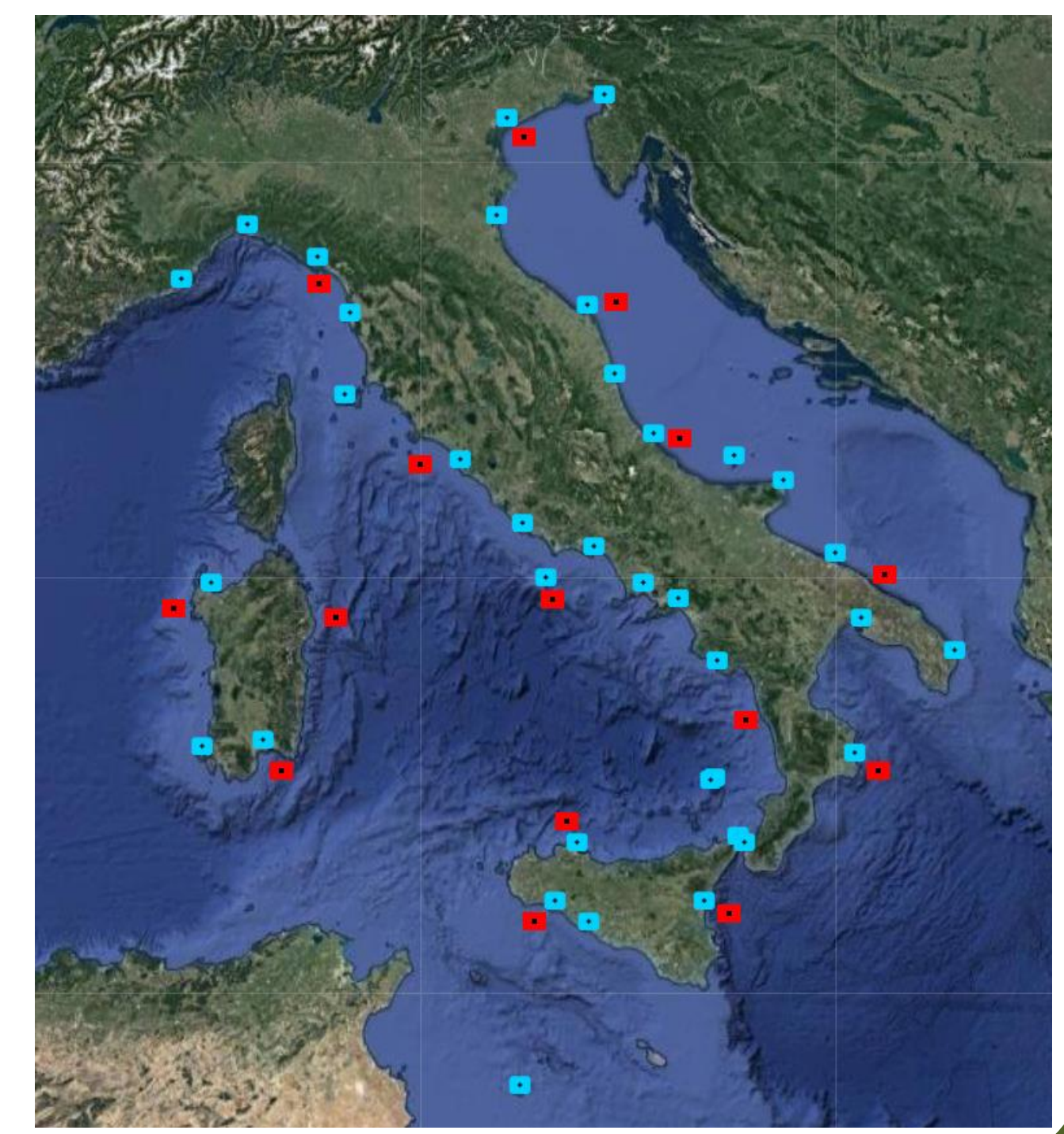
Linked ISPRA is an prototype project of ISPRA (Italian Institute for Environmental Protection and Research), started in 2016 based on the development of an application, according to the World Wide Web Consortium (W3C) specifications to produce and publish LOD. To be compliant with national and international legislation on digital public administration, since the beginning the application is developed with a data-oriented approach rather the service-oriented systems. The Linked ISPRA platform provides access to several national environmental datasets, under the terms of the CC-BY 4.0 license. The technology stack is completely developed with open source components.



## MARINE MONITORING NETWORKS

*A sea of data*

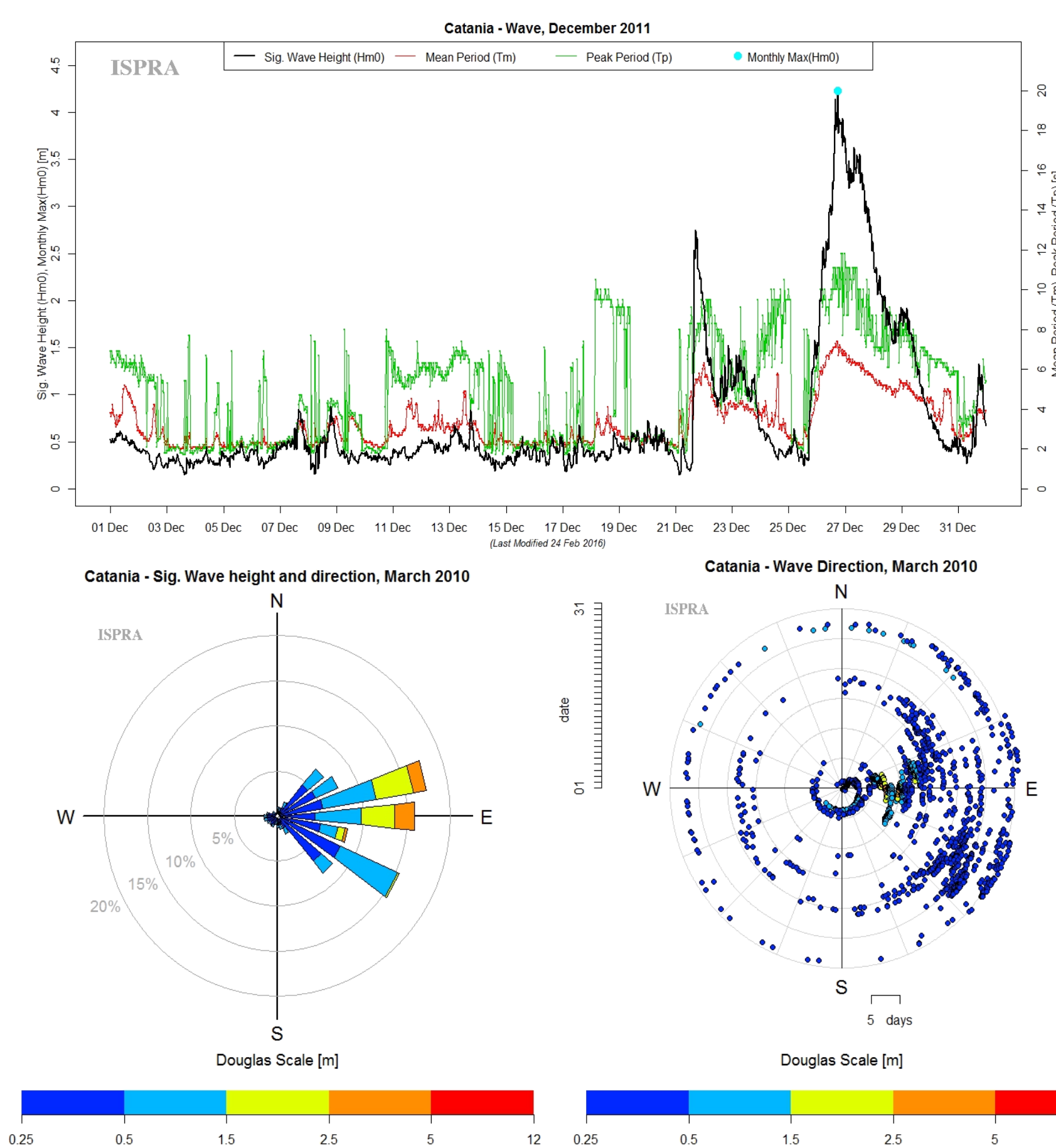
ISPRA directly manages two national marine monitoring networks: the Italian data buoy network (RON – Rete Ondametrica Nazionale) and the Italian national tide gauge network (RMN – Rete Mareografica Nazionale). RON includes 15 directional wave buoys (red dots) and the corresponding land based receiving stations collecting real time series of wave and meteorological parameters every 30 minutes. It provides data needed to define the sea state, such as: significant wave height, peak period, mean period, wave direction, sea surface temperature, wind speed and direction, air temperature, atmospheric pressure, relative humidity. RMN includes 36 coastal stations (blue dots). It is the hugest network for sea level measurement in the Mediterranean, collecting real time series of sea level and meteorological parameters every 1 minute. It integrates the National meteorological system. Data gathered are used for determining tides and seiches and can support the identification and characterization of early tidal waves (tsunami) and of storm tides. Both networks work almost continuously from eighties and data are published in real time (RON is suspended for maintenance at the moment). Relative exposed metadata consist by a set of information such as the type of parameters, the measurement period, the used instrument, the validation level, expressed as L0 (on-flight data), L1 (automatic validation), L2 (expert judgment).



## EXAMPLE

SPARQL query for wave data in Catania (Sicily) from 2009 to 2014 on <http://dati.isprambiente.it/sparql>

```
select distinct ?period ?csvUrl where {
  ?parameter a :Wave.
  ?place rdfs:label «Catania».
  FILTER (str(?period)>='2009-01' AND str(?period)<='2014-12').
  ?parameter gn:nearbyFeature ?place.
  ?collection a :MeasurementCollection;
  :measurementPeriod ?period;
  :isDataOf ?parameter;
  dcat:downloadURL ?csvUrl.
} ORDER BY ?period
```



## SEMANTIC WEB: ONTOLOGIES, VOCABULARIES & THESAURI

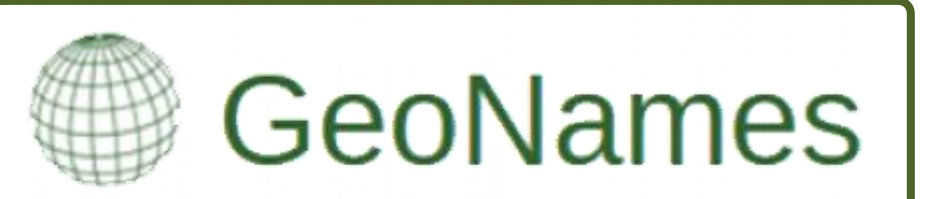
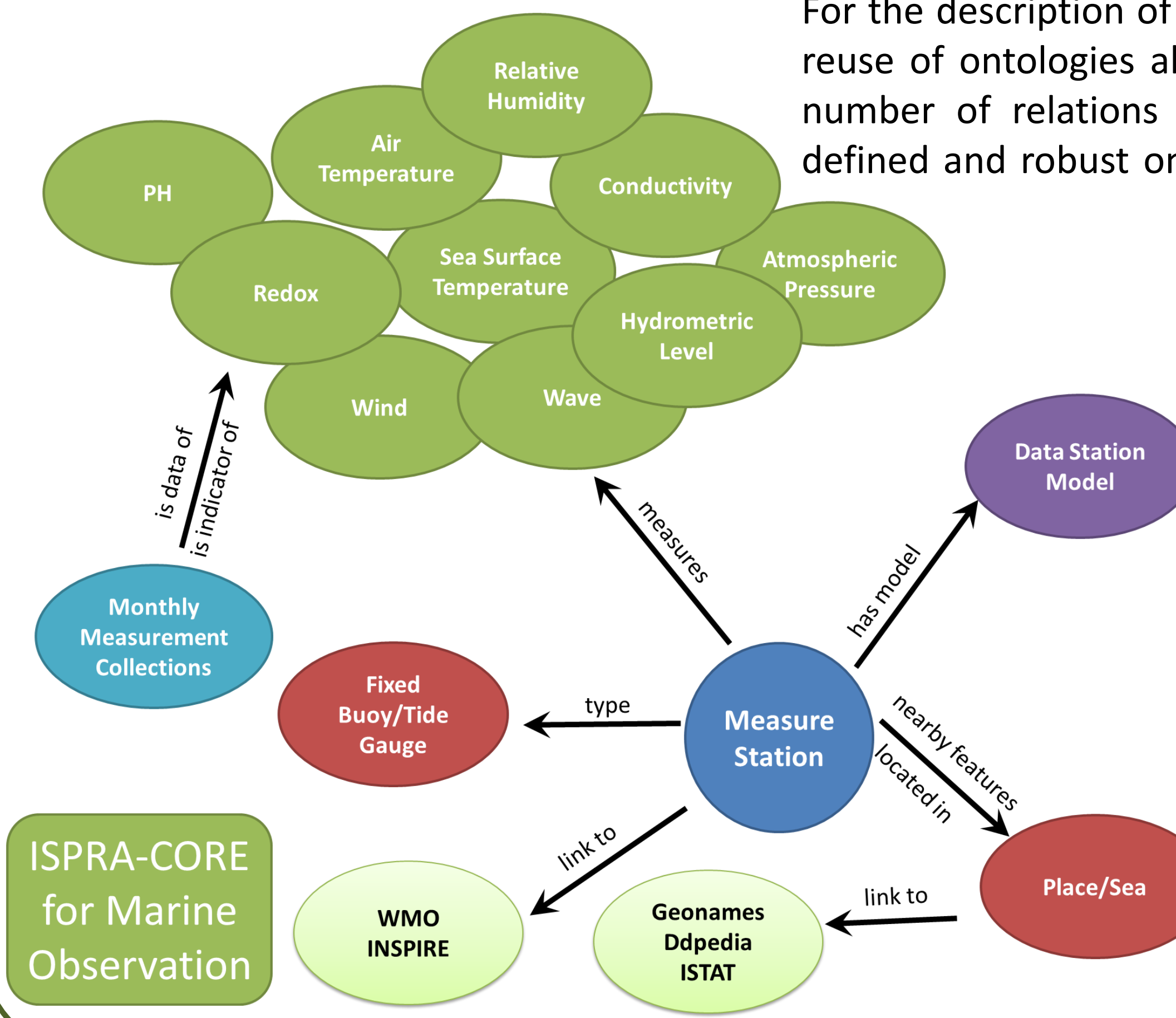
*"I have a dream for the Web, become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A Semantic Web, which makes this possible..." (Tim Berners-Lee)*

The Semantic Web is an extension of the ordinary Web and provides a common framework that allows data to be shared and reused from humans and machines across application, enterprise, and community boundaries. This data management is strongly enhanced by the application of knowledge representation languages that use concepts, classes and relations that are typically modeled throughout ontologies, vocabularies and thesauri.

For the description of generic and most common metadata, the direct reuse of ontologies already existing is deeply recommended. A great number of relations among classes are defined starting from well defined and robust ontologies such as "Dublin Core", "FOAF", "SKOS", "Schema", "Geonames", "QUDT".

In particular for marine dataset, the World Meteorological Organization Code Registry (<https://codes.wmo.int>) is one of the most useful and complete concepts repository in order to describe marine monitoring networks and observations.

On the other hand, the needs of classes and properties for the specific dataset domain bring to the creation of suitable and widely documented resources. ISPRA-CORE is the new ontology of the Institute born to supply the lack of accuracy of generic existing ontologies in order to define marine measurements, indicators and properties relates to marine monitoring networks.



## BEST PRACTICES

*Can machine understand human stuff?*

- Follow the guidelines: Italian and European communities and agencies suggest good practices to ensure semantic interoperability (see last releases of DCAT-AP and GeoDCAT-AP for datasets description);
- Each resource is defined and documented through an URI (an address on the World Wide Web) that must be persistent in time;
- Reuse existing and robust ontologies avoiding the proliferation of identical information in multiple vocabularies;
- Resources is not only for machines, use a human-readable information;
- Spread the resources through links with other existing resources domains;
- Good resources, enlarged metadata, Big Data.

## OUTPUT

RDF/XML  
N-Triples  
CSV  
JSON  
HTML  
...

## LINKED ISPRA IN NUMBERS

|   |           |
|---|-----------|
| Dataset already published                 | 5         |
| Total number of triples                   | 3.796.525 |
| Triples related to marine resources       | 254.597   |
| Monthly marine data/indicator collections | 17.871    |
| Marine measurement stations               | 76        |
| Active users per day                      | About 500 |