Data e-Infrastructure Initiative for Fisheries management and Conservation of Marine Living Resources

The iMarine Data Bonanza

Improving Data Discovery and Management through an **Hybrid Data Infrastructure**

Donatella Castelli, **Pasquale Pagano**, Leonardo Candela, Gianpaolo Coro
iMarine Technical Director
pasquale.pagano@isti.cnr.it
ISTI – CNR (Italy)

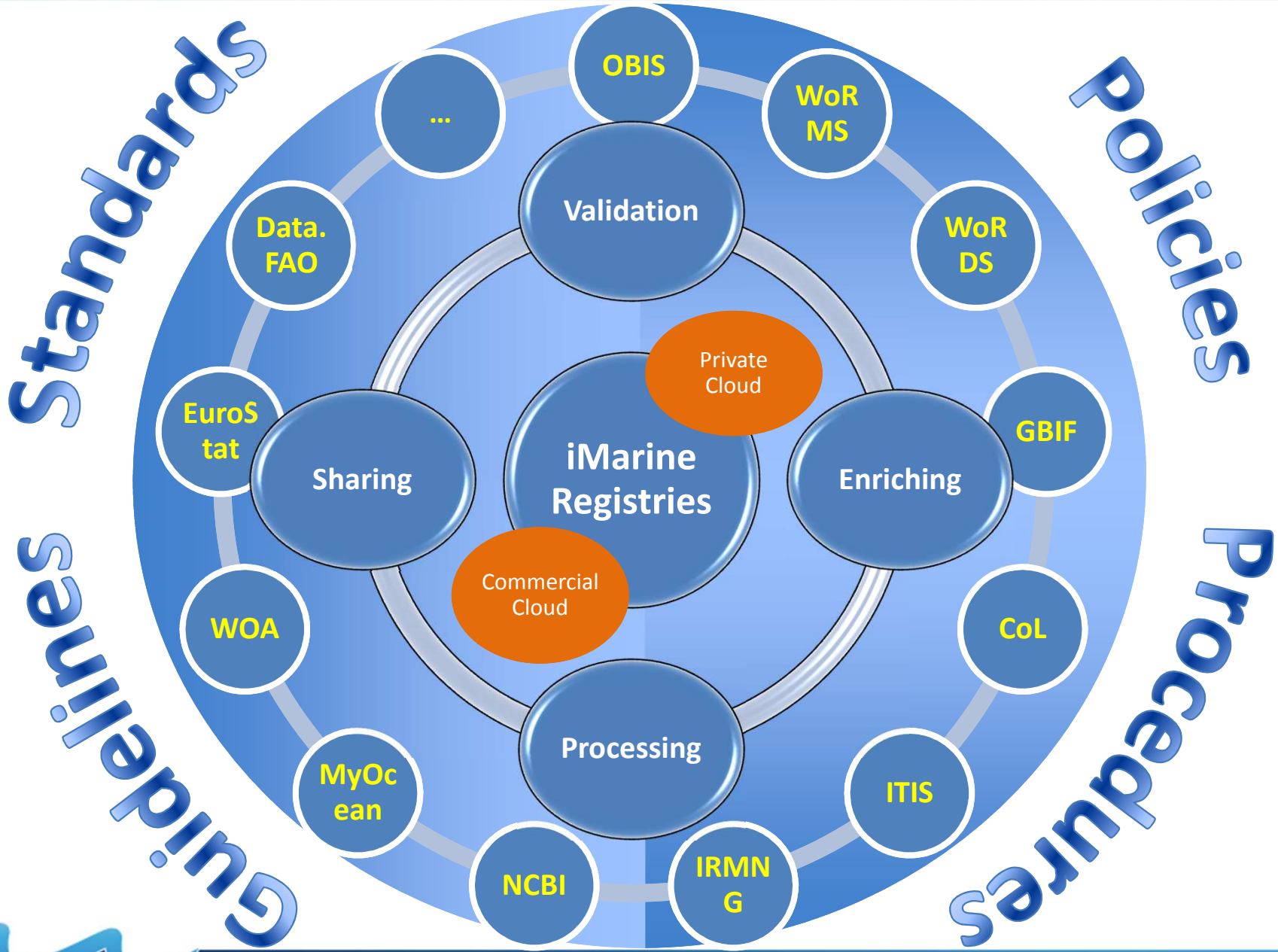
i-marine.eu supports the principles of the Ecosystem Approach (EA) to fishery management and conservation of marine living resources

- EA is a strategy for the integrated management of land, water and living resources that promotes conservation and sustainable use in an equitable way

i-marine.eu operates an Hybrid Data Infrastructure (HDI) offering access to a rich array of marine-related data and products via tailored environment

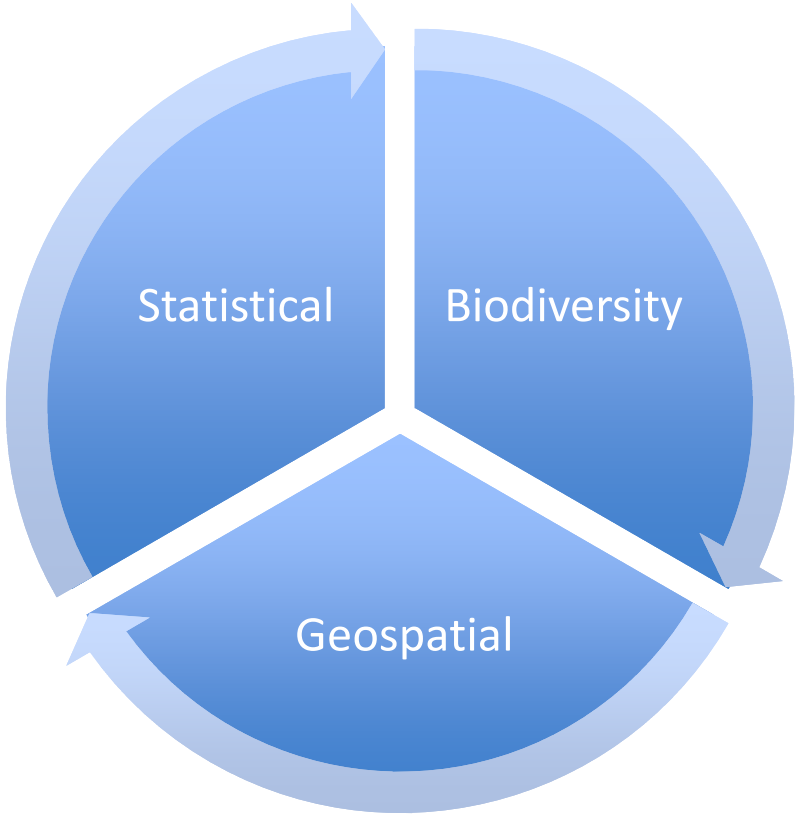
- HDI implements a data-management-capability delivery model in which computing, storage, data and software are made available as a utility *(as-a-Service)*

e-Infrastructures Ecosystem



Data Bonanza

SDMX *
- FAO CodeLists
- IRD CodeLists
- FAO Global Aquaculture Production
- FAO Global Capture Production
- FAO Global Production
- Eurostat
- …



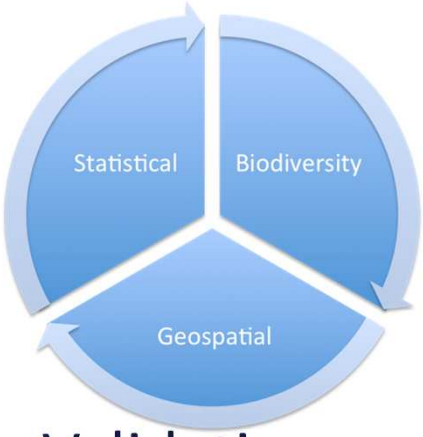
DarwinCore / ISO19139
>35 M Observations (OBIS)
≈ 120 K Observed Species (OBIS)
≈ 500 K Taxa (WoRMS)
>600 K Scientific Names (ITIS)
>12 K Species Distribution Maps (AquaMaps)
≈ 600 Species Extent (FAO)

… FishBase, SeaLifeBase
… CoL, GBIF

> 300 variables

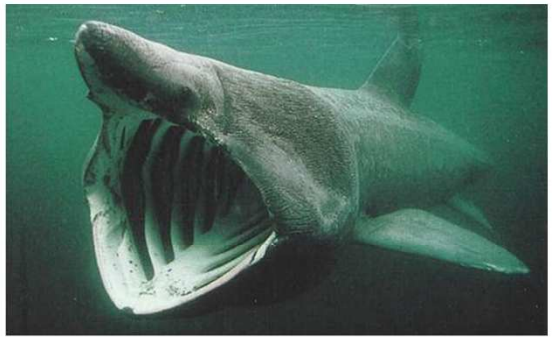
ISO19139 (OGC W*S)
- 10 years Chemical and Physical variables in 2D space
  - Ice concentration and velocity, Chlorophyll, Oxygen, Nitrate, Phosphate, Phytoplankton as carbon, Salinity, Temperature, …
- On-demand Chemical and Physical variables in 3D space
  - Apparent Oxygen Utilization, Dissolved Oxygen, Salinity, Temperature, …

Not Only Access

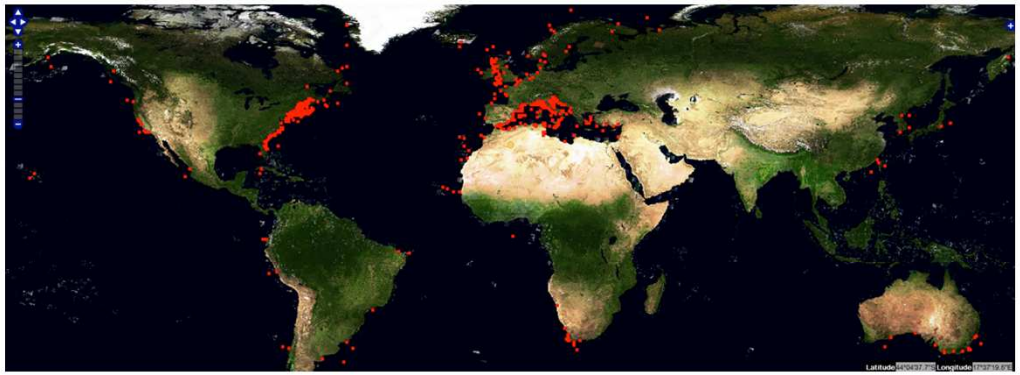


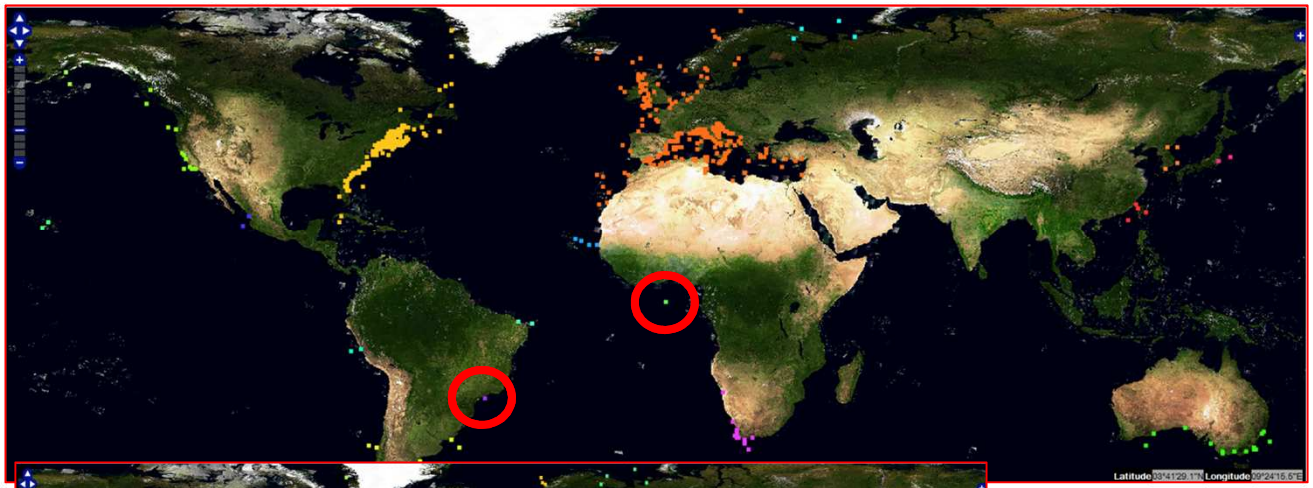
- Access
  - Retrieval of geospatial data as space/time-varying phenomena
  - Direct fine-grained access to feature and feature property level.
- Validation
  - User-defined quality and dissemination level
- Enriching
  - Generation metadata, exploitation of reference data, linking to environmental dataset
- Processing
  - Analysis and mining exploiting e.g. R, Weka and RapidMiner statistical frameworks
- Sharing
  - User-driven process to decide how other agents (human / machine) can access information

Features Clustering with **StatsCube**



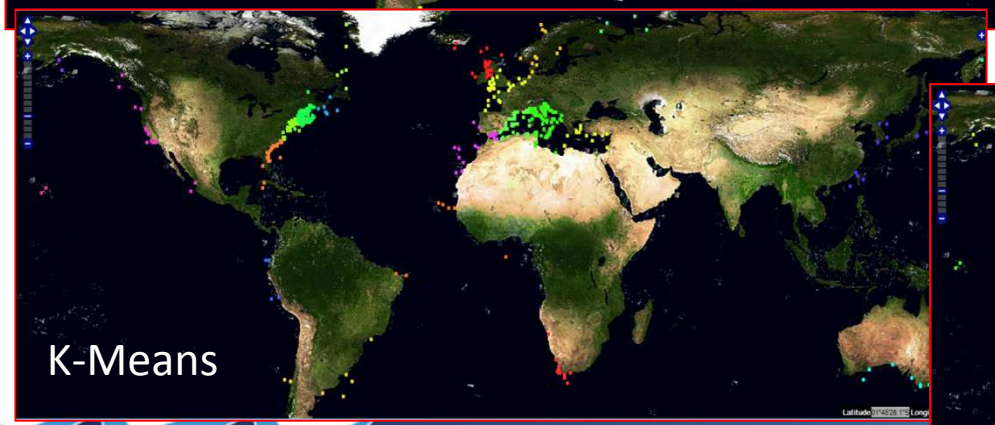
Presence
Points
(FishBase
+
Obis)

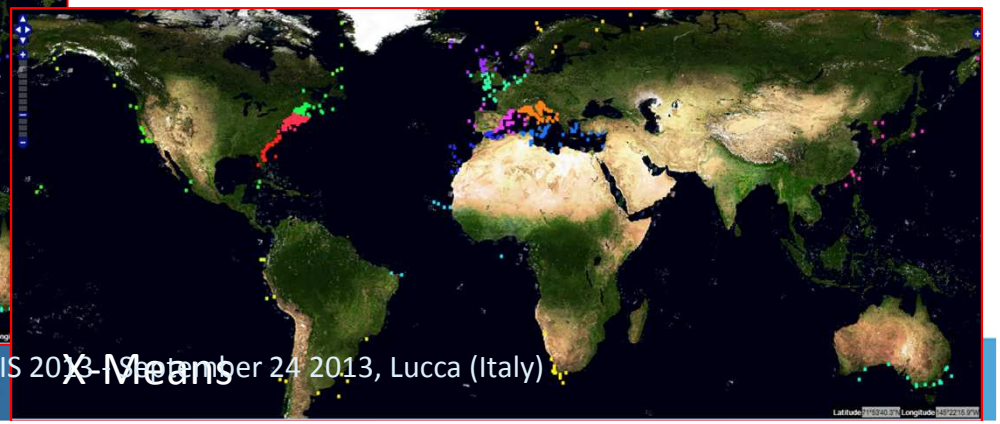




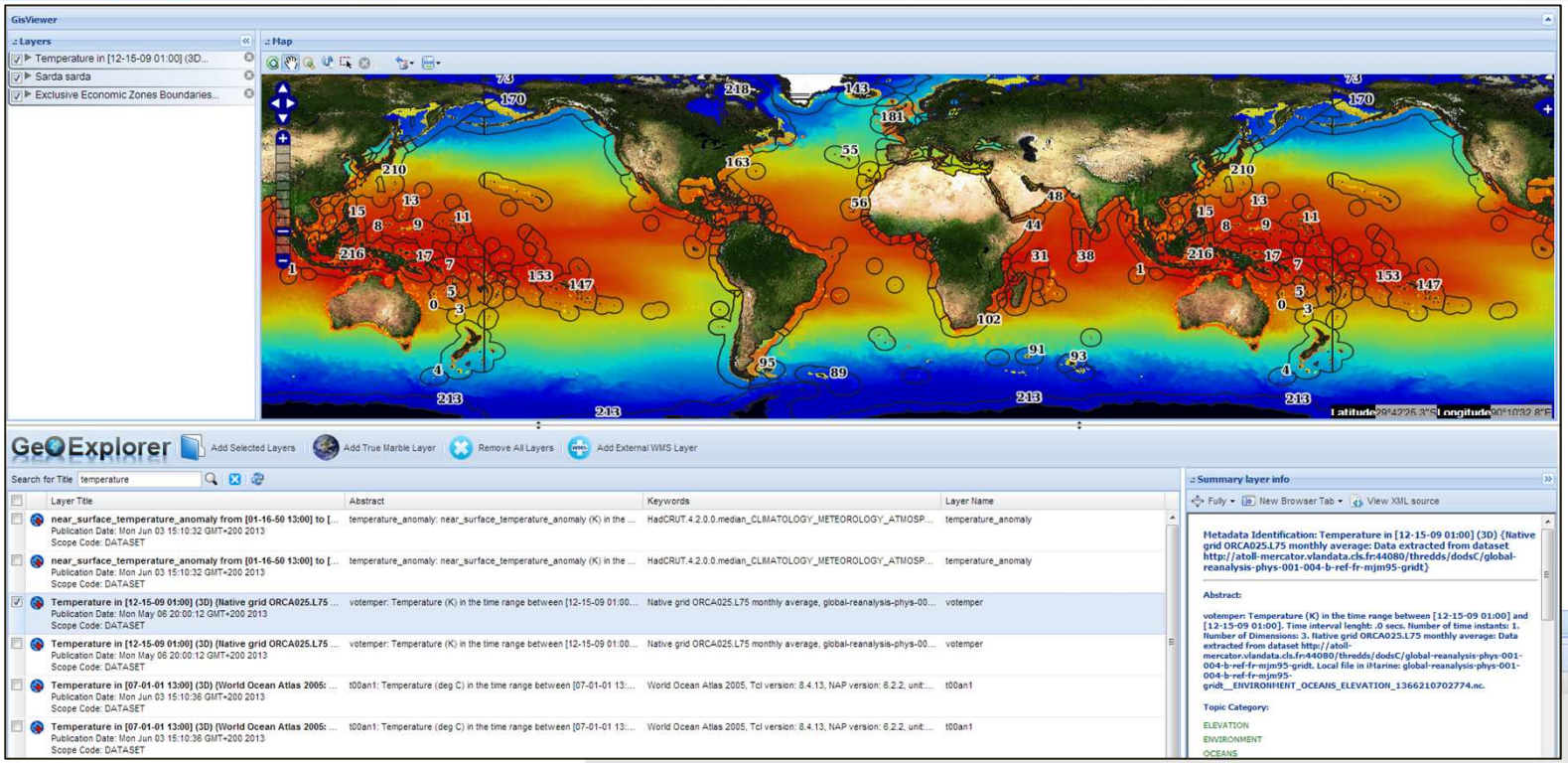
Density Based Clustering
DBSCAN
(with outliers)

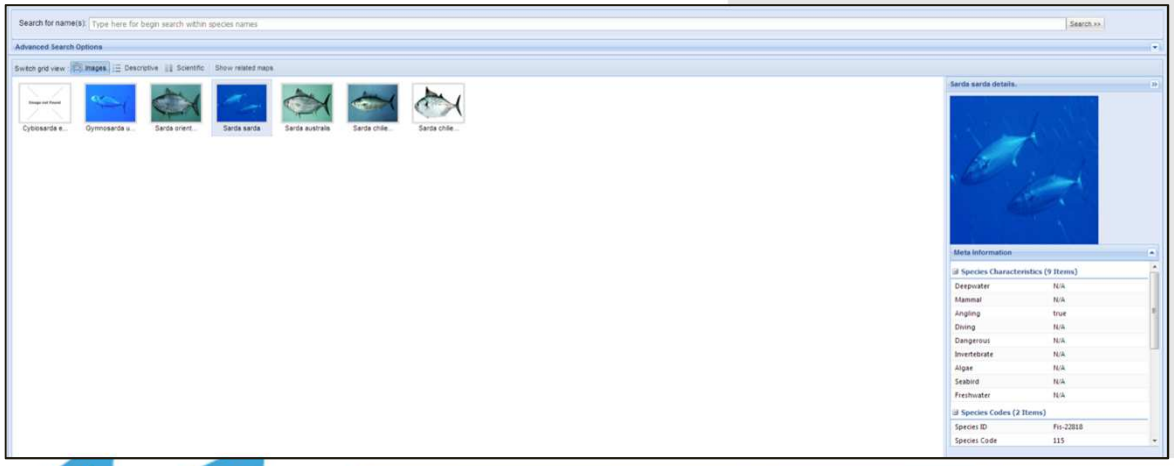
Other methods are also available …

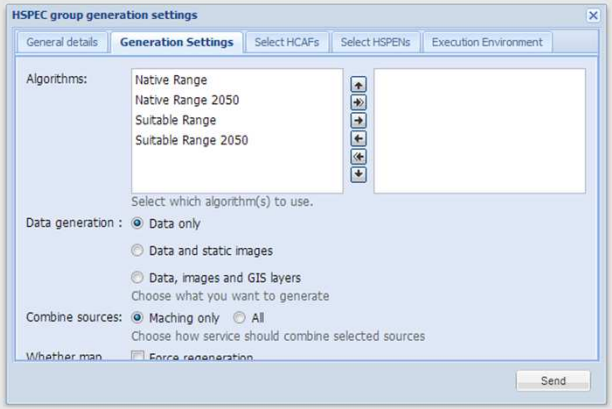




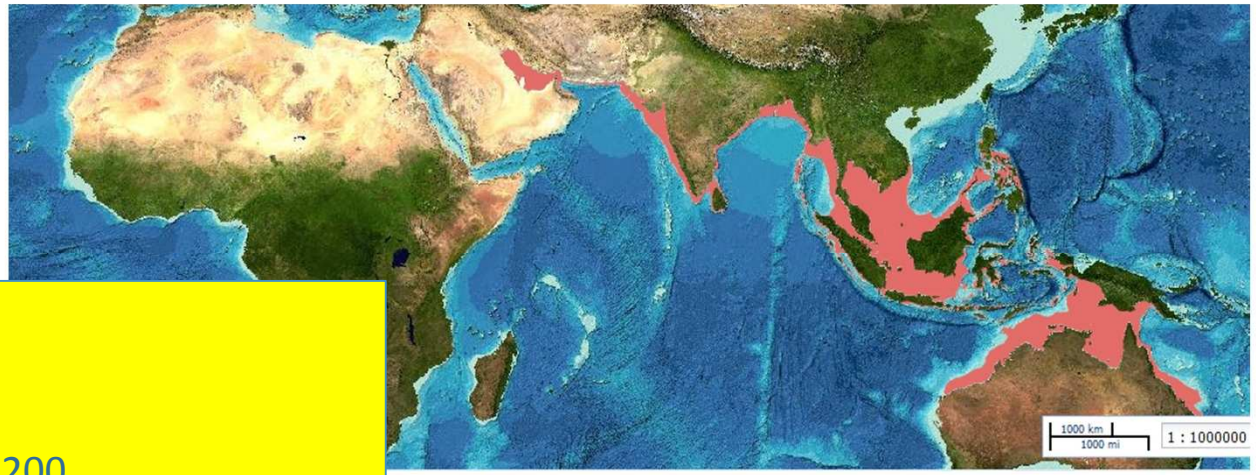
Ecological Modelling with **BiolCube**







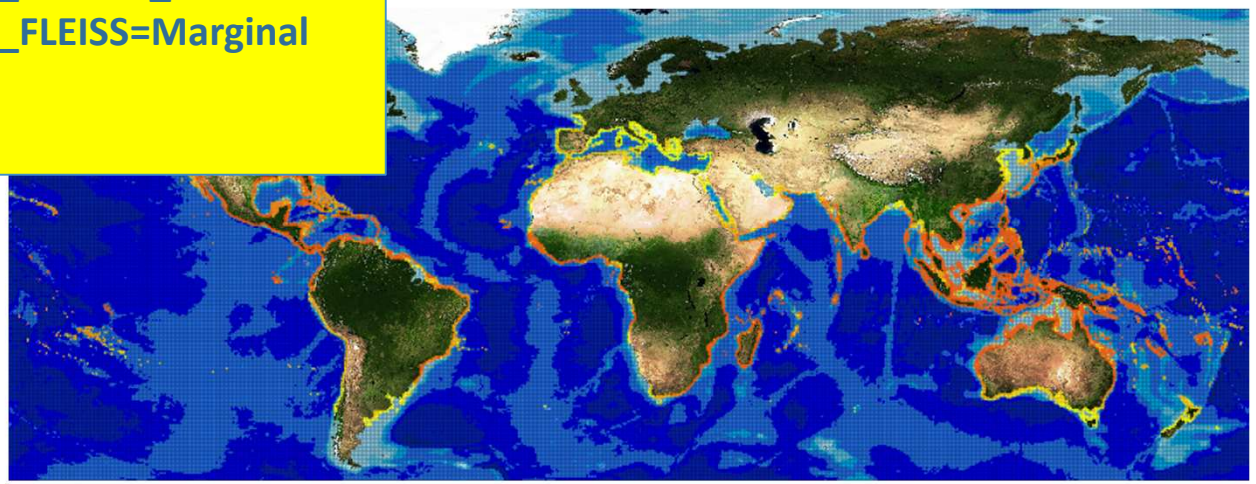
Maps Comparison with **GeosCube**



FAO Eleutheronema tetradactylum

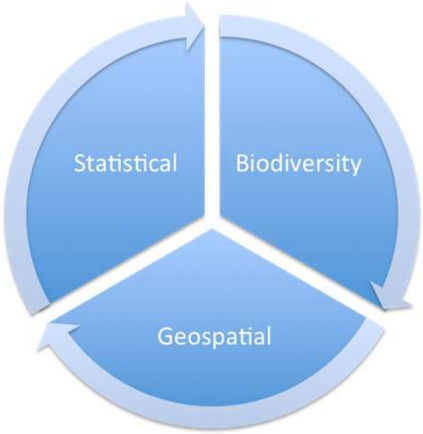
VS

AquaMaps Eleutheronema tetradactylum



MEAN=0.81
VARIANCE=0.02
NUMBER_OF_ERRORS=6691
NUMBER_OF_COMPARISONS=259200
ACCURACY=97.42
MAXIMUM_ERROR=1.0
MAXIMUM_ERROR_POINT=3005:363:1
COHENS_KAPPA=0.218
COHENS_KAPPA_CLASSIFICATION_LANDIS_KOCH=Fair
COHENS_KAPPA_CLASSIFICATION_FLEISS=Marginal
TREND=EXPANSION
RESOLUTION=0.5

Not Only Access, Validation, Enriching, Processing, Sharing



- It is always possible to **save** the discovered data in various Standard formats
- It is always possible to **collaborate** with co-workers through a dedicated workspace.

- **Mash-up** data across diversity
  - Accessing statistical datasets in SDMX, geo-referencing them, describing them in ISO19139, and making them available via OGC W*S standard protocols
  - Accessing species observation datasets in DwC, analysing their distribution trend via R, and projecting them in geographical space
  - Accessing species taxonomies in DwCA and publishing them as reference data in SDMX

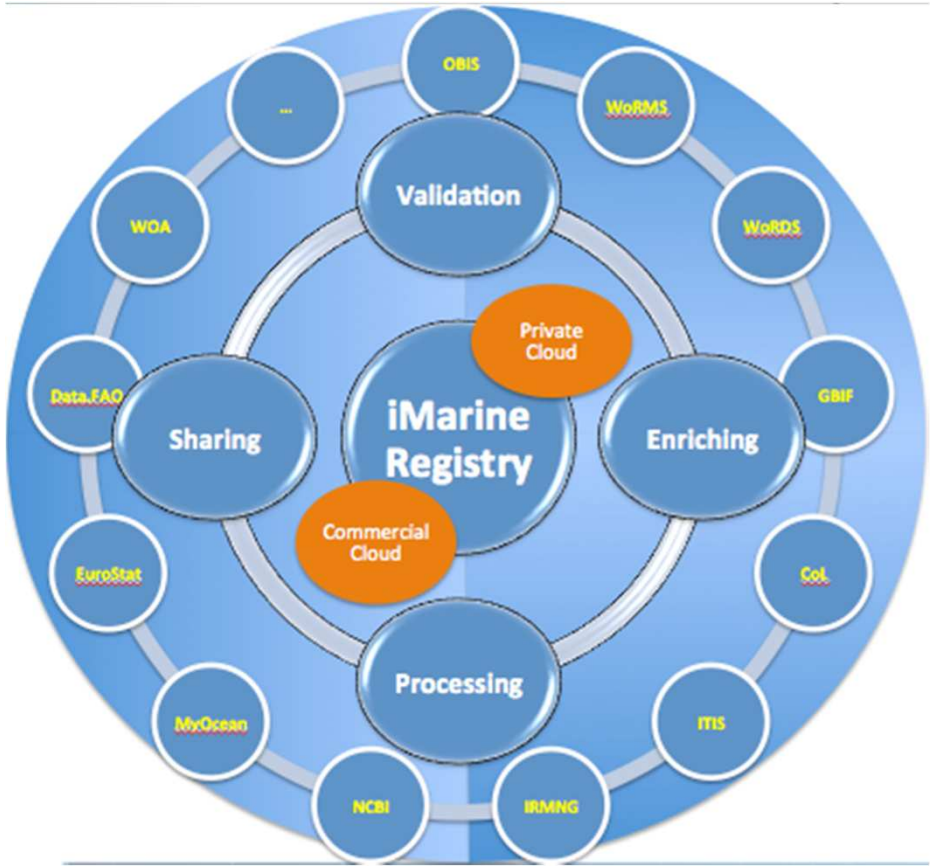
iMarine Data Policies

- The data and information are available under terms described in the product metadata

- Except where otherwise noted, this is the **Creative Commons License**

- All derivative products are licensed under the **Creative Commons License CC BY-SA**

Is this enough?



- An **ecosystem** of participatory data e-Infrastructures
- Regulated by **policies**
- Enabled by **standards**
- Promoting not only access but **mash-up** of heterogeneous data

User centric

User-Centric View

User-centric view of an **ecosystem** of participatory data e-Infrastructures to

- Cope with the overwhelming amount of data and capacities
- Promote re-use of data
- Encourage sharing of resulting products

User-centric and workflow-oriented

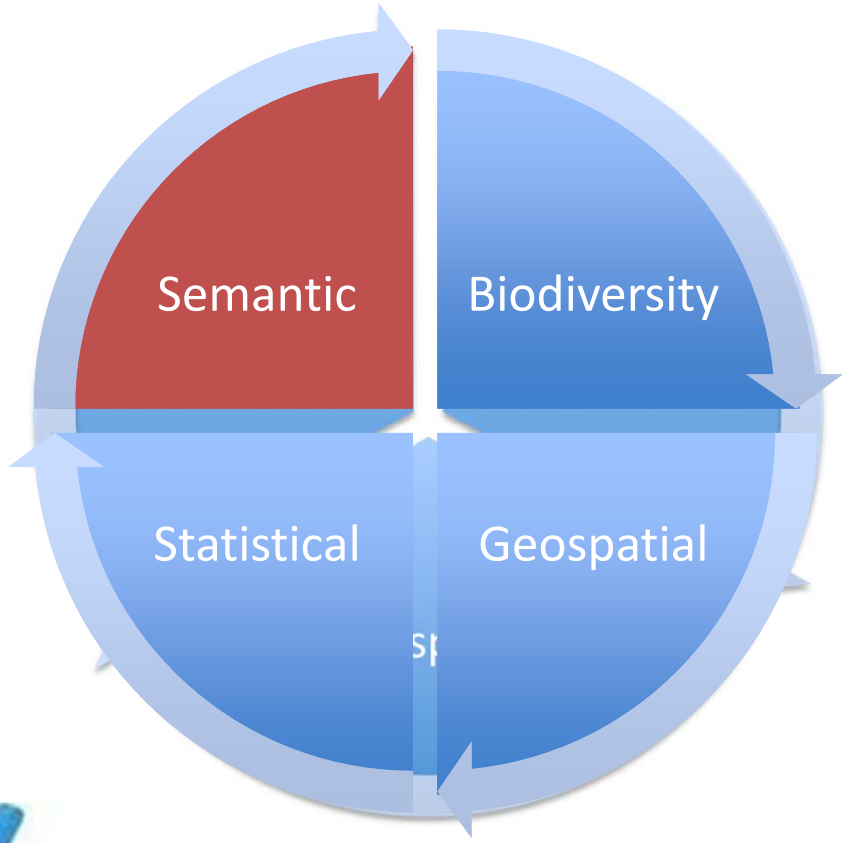
Virtual Research Environment

iMarine is user-centric and workflow-oriented thanks to the gCube VRE technology

Virtual Research Environment (VRE) is

- a **distributed and dynamically created** environment
- where **subset of** data, services, computational, and storage **resources**
- regulated by **tailored policies**
- are **assigned to a subset of users** via interfaces
- for a **limited timeframe**
- at **little or no cost** for the providers of the participatory data e-infrastructures

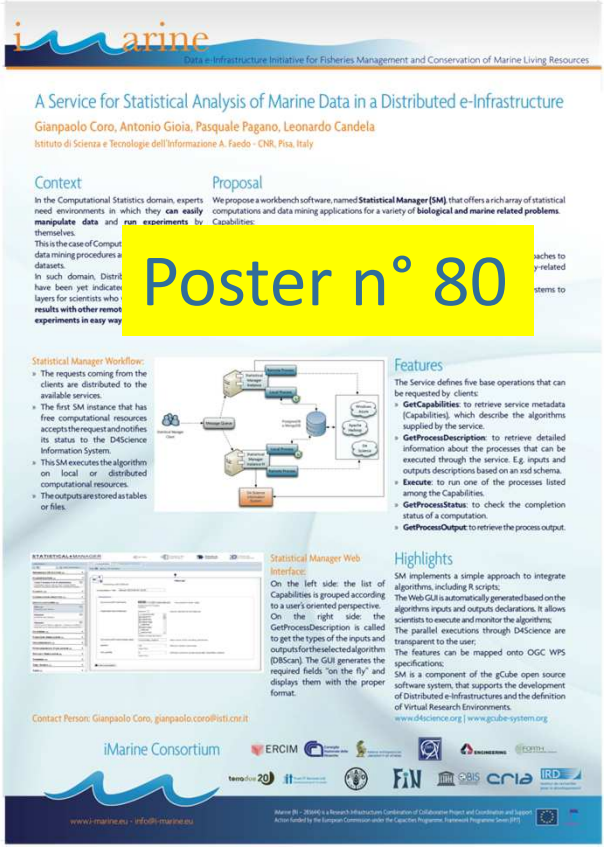
Conclusions



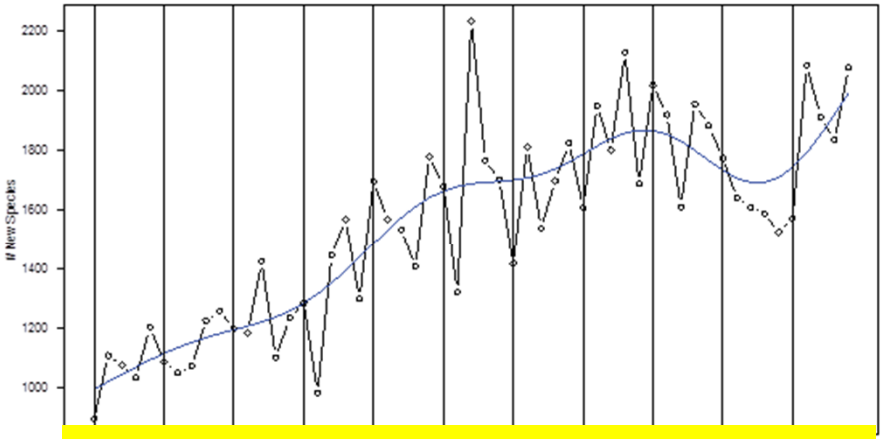
More at IMDIS 2013

A Service for Statistical Analysis of Marine Data in a Distributed e-Infrastructure

Gianpaolo Coro, Antonio Gioia, Pasquale Pagano, Leonardo Candela



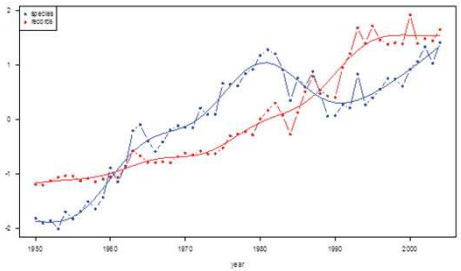
Poster n° 80



Wednesday 25, September 2013 – 9:40

Trendylyzer: a Long-Term Trend Analysis on Biogeographic Data

***Ward Appeltans**, Peter Pissierssens, IOC-UNESCO*
Gianpaolo Coro, Angela Italiano, Pasquale Pagano, ISTI-CNR
Anton Ellenbroek, FAO
Tom Webb, University of Sheffield



Thanks for your attention

Visit

https://i-marine.d4science.org

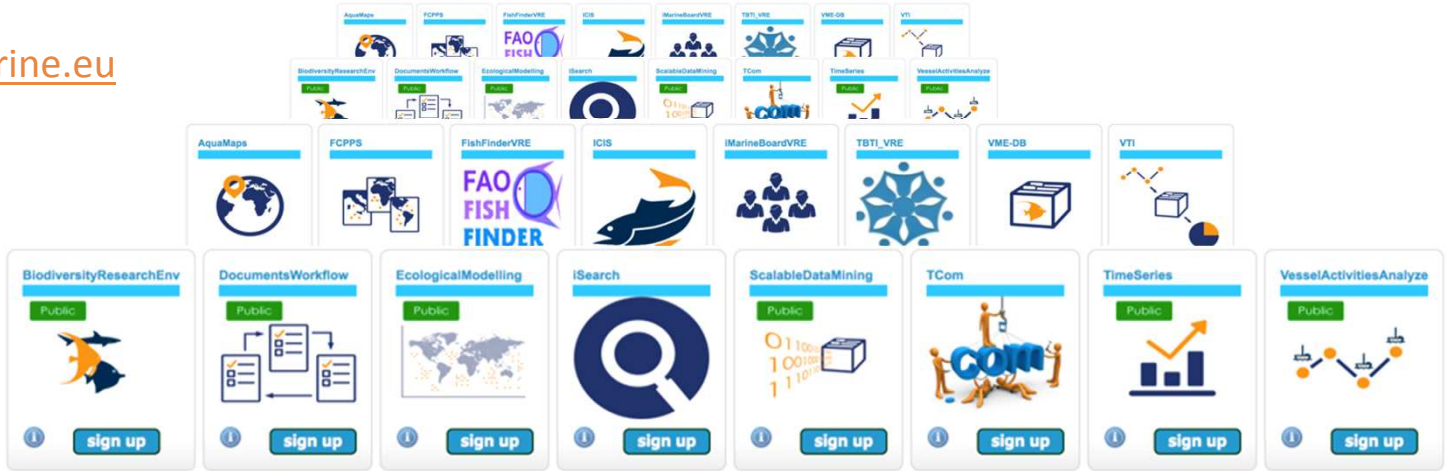
http://www.i-marine.eu

Join

gCube Apps

Enjoy

applications



gCube Application Bundles

gCube is the iMarine empowering technology

BIOL CUBE — Management and interpretation of biological and ecological data in the environment

STATS CUBE — Complete full life-cycle data framework, from observational data to aggregated data repositories enriched with validation and analytical tools

GEOS CUBE — Storage and interpretation of geospatial explicit information, including WPS processing

CONNECT CUBE — Flexible sharing, storage, reporting, search and retrieval, aggregation and projection facilities

A BUNDLE is a set of services and technologies grouped according to a family of related tasks for achieving a common objective