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

iMarine Overview

Donatella Castelli
CNR-ISTI

Outline

- iMarine Project & e-Infrastructure Overview
- gCube Enabling Technology
- Useful Links

Project information

EU FP7 Capacity e-Infrastructure project, Nov 2012- Apr 2014




Finance and admini
- ERCIM

Dissemination
- Trust-IT

Technology
- CNR
- NKUA
- CERN
- E-IIS
- FORTH
- TERRA2

Members of the Community
- CRIA
- FAO
- FIN
- IRD
- UNESCO
- *NEAFC*

www.i-marine.eu

iMarine Board

i-marine Board

NEAFC, IUCN, Wageningen Univ., DG-MARE, MEDDE, Univ. of Amsterdam, UKMarine Mgmt Organization, VLIZ, ICES, ESTAT, IODE-UNESCO, FAO, FIN, CRIA, IRD, TerraDue

Extended

ATZI-Tecnalia
IEO (Spanish Institute of Oceanography)
SFPA (Sea-Fisheries Protection Authority)
SFP (Sustainable Fisheries Organisation)
CI (Conservation International)
SEAFO (South East Atlantic Fisheries Org.)
TBTI (Too Big To Ignore)
NAFO (Northwest Atlantic Fisheries Organisation)
IOTC (Indian Ocean Tuna Commission)
GRID-Arendal
Science Europe
Agrocampus OUEST
MEDPAN (Network of Marine Protected Area Managers in the Mediterranean)
CCAMLR (Commission for the Conservation of Antarctic Marine Living Resources)
UNEP WCMC (World Conservation Monitoring Centre)
FAO FIRF (Marine and Inland Fisheries Service)
ENEA (Italian National Agency for New Technologies, Energy and Sustainable Economic Development)

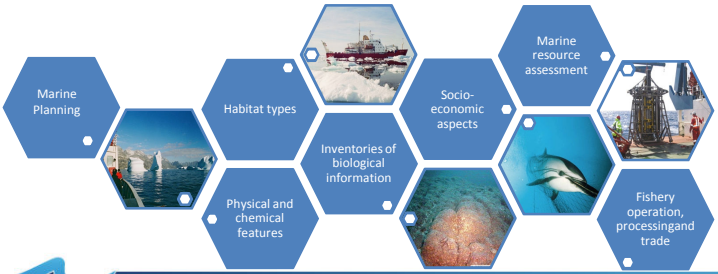
iMarine Objective

Launch an Initiative aimed at establishing and operating a data Infrastructure supporting the implementation of the principles of the **Ecosystem Approach to Fisheries Management and Conservation of Marine Living Resources**

Challenges

Capacity & Diversity

- Analysis and processing of a large amount of **heterogeneous, across-domain produced information**
- **Multidisciplinary & multifacets collaboration at the local**, national, regional and international levels



e-Infrastructure

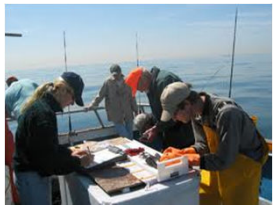
Elecronic platform operated by a **responsible entity** offering an open set of **basic enabling facilities** (including access to resources) "as-a-service" to a distributed **Community of Practice**.

Community of Practice point of view: *Services*

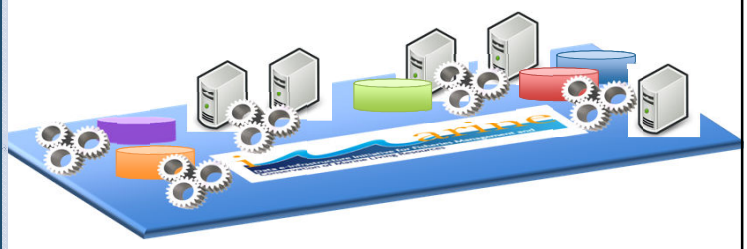
Services

- Collaboration
- Sharing
- Visualization
- Publication
- Storage
- Processing
- Analysis
- Modeling
- Integration
- Enrichment
- Transformation
- Validation
- Access
- Discovery

From science to policy making



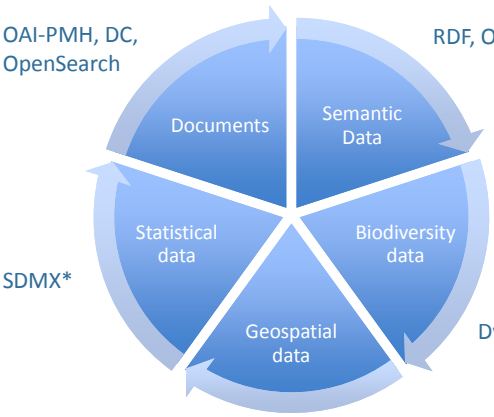
- **Analysis and processing** of a large amount of heterogeneous, across-domain produced information
- **Multidisciplinary & multifacets collaboration**



Across hetogeneous data

Services

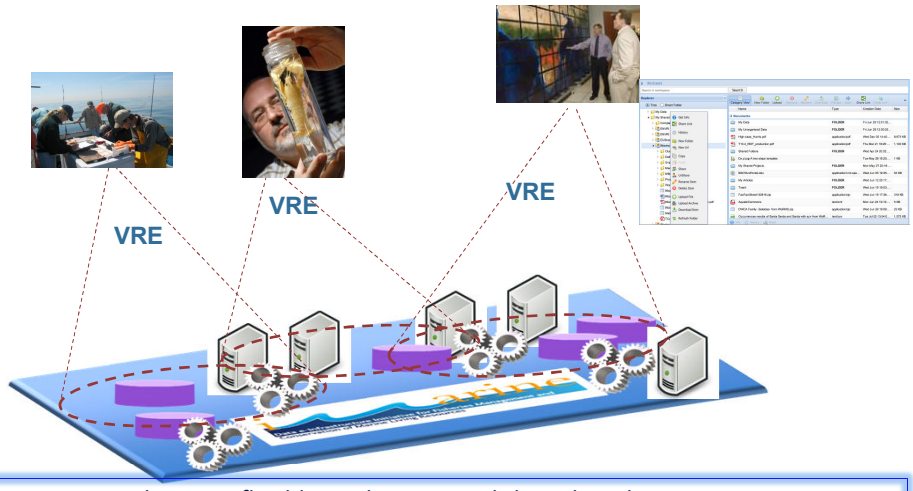
- Collaboration
- Sharing
- Publication
- Storage
- Processing
- Analysis
- Modeling
- Integration
- Enrichment
- Transformation
- Access
- Discovery



Community of Practice point of view:

Exploiting service through APIs

Community of Practice point of view:
Virtual Research Environments



VRE = Comprehensive, flexible, and secure Web-based working environment serving the application needs of a community working for a specific goal

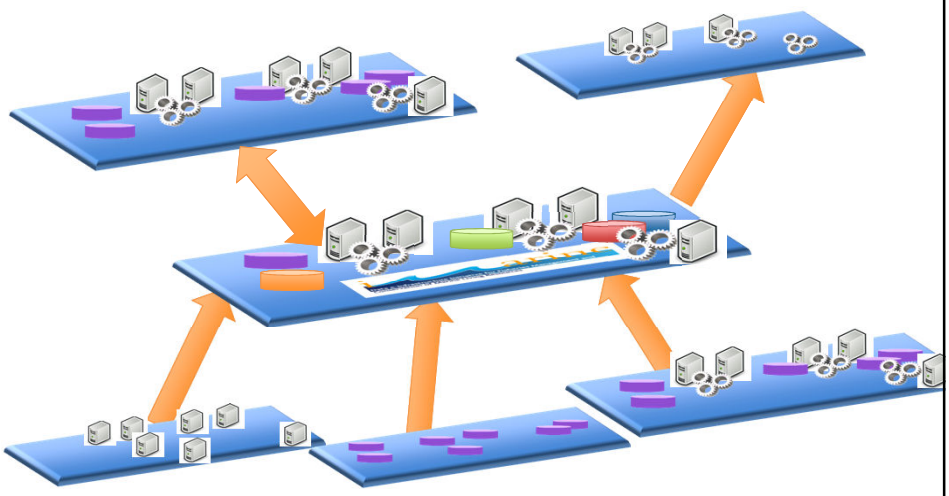
iMarine VREs

Data Management VRE Examples

VRE	Purpose	Target
ICIS	Facilitates the collection, curation and dissemination of Fisheries statistics	Fishery Statisticians, Marine Biologists
Time Series	Facilitates the collection, curation and dissemination of Statistical data.	Statisticians
Vessels Activities Analyzer	Performs Data Mining on Vessel trajectories.	Fishery Statisticians
Scalable Data Mining	Applies Data Mining techniques to biological data.	Statisticians
AquaMaps	Predicts and produce species presence maps	Fisheries and Aquaculture Scientists

e-Infastructure ecosystem

building on previous efforts



Inside iMarine: *gCube*

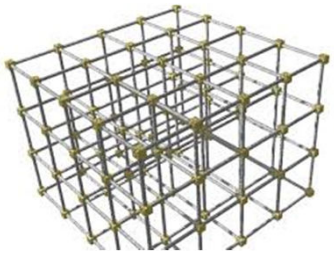
gCube Framework

385 components
(web services, libraries, portlets)

Hosted Technology and data

- Cassandra/Hadoop/MongoDB Clusters
- Virtuoso TripleStore
- SDMX registry
- OGC services cluster (GeoNetwork, Geoserver, WPS, Thredds)
- FishBase and SeaLifeBase DBs
- .....

gCube frameworks



gCube Framework

Facilitated development of new tools

Exploitation of standards

New tools can become infrastructure resources available according certain policies

Enabling services

- Deployment
- Hosting
- Resources Lifecycle
- Monitoring
- Accounting
- Security
- Distributed Workflow Execution

iMarine Product Catalogue

Application Bundles

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 environment for semantic resources creation and exploitation

A BUNDLE is a set of services and technologies for supporting related tasks

BIOL CUBE

A subset of the products and services belonging to BiolCube

https://gcube.wiki.gcube-system.org/gcube/index.php/Biodiversity_Access

Species Data Discovery

Search: Taxon By: Scientific name Term: e.g. sarda sarda, solea solea Search Example

Search for multiple species

Advanced Option | Filter by Source | Filter by BBox | Filter by Date | Synonyms From | Expand

Occurences Sources: GBIF OBIS

Classification Sources: CatalogueOfLife IRMNG ITIS NCBI OBIS WoRDSS WoRMS

Search across several data providers

Search in GBIF all the occurrences about 'sarda sarda' and its synonyms found in WoRMS

Search in OBIS all the occurrences for 'sarda sarda' and 'Carcharodon carcharias' expanded with synonyms from WoRMS and CoL. Accept only the results with an event date between 2000 and 2005.

Occurrence Points

Occurrence Data from **GBIF**

Occurrence Data from **Obis**

∩	U	-	DD
Intersection	Union	Difference	Duplicates Deletion

A
x,y
Event Date
Modif Date
Author
Species Scientific Name

Records

Similarity

B
x,y
Event Date
Modif Date
Author
Species Scientific Name

Similarity between habitats

Habitat Representativeness Score:

1. Measures the **similarity** between the environmental features of two areas
2. Assesses the quality of models and environmental features

Latimeria chalumnae

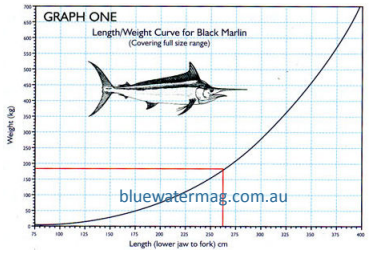
HRS=10.5

Habitat Representativeness Score

Length-Weight Relationships

$$W = aL^b$$

Objective:
Calculate the *a* and *b* parameters for 'several' species.



Approach:
- Collaborative development with GEOMAR
- Integration of R Scripts
- Usage of Cloud computing for R Scripts

- The scripts are run in distributed fashion
- The time of the scientist procedure was 20 days
- The time on the Statistical Manager was 11 hours! **Time reduction of 95.4%**
- The script has been run periodically and currently solves LWR for **37 234 species**

STATS CUBE

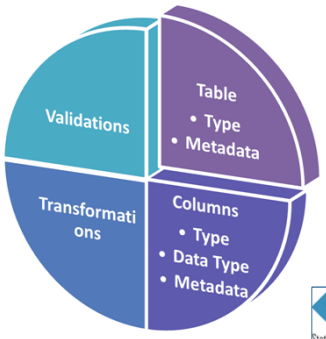
A subset of the products and services belonging to StatsCube

http://wiki.i-marine.eu/index.php/Catalogue:Applications#StatsCube

Data lifecycle

- Import
- Store
- Transform
- Validate
- Harmonize
- Analyze
- Visualise
- Publish

Dataset compliant with a specified template



A table template defines:

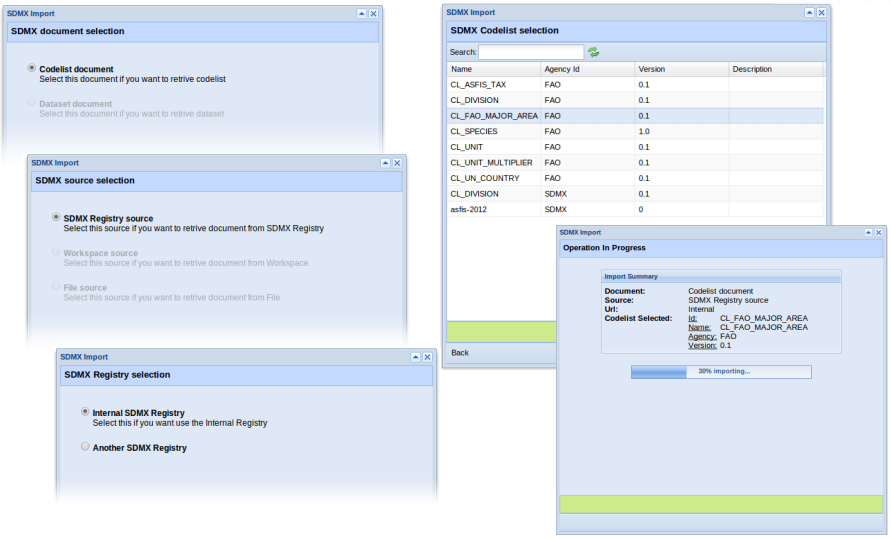
- Table definition, columns definition, set of table transformations, set of validation procedures

Can be applied to any dataset

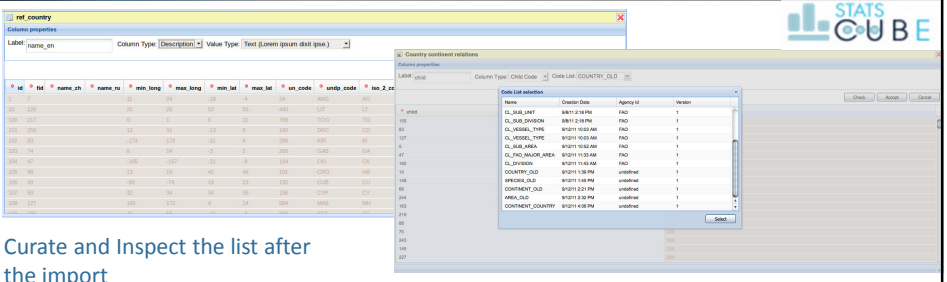
Can be modified and shared

20th Marine Observation and Data Expert Group Meeting, 21-22 October 2013, Brussels

SDMX Code list import wizard



Code list Manager

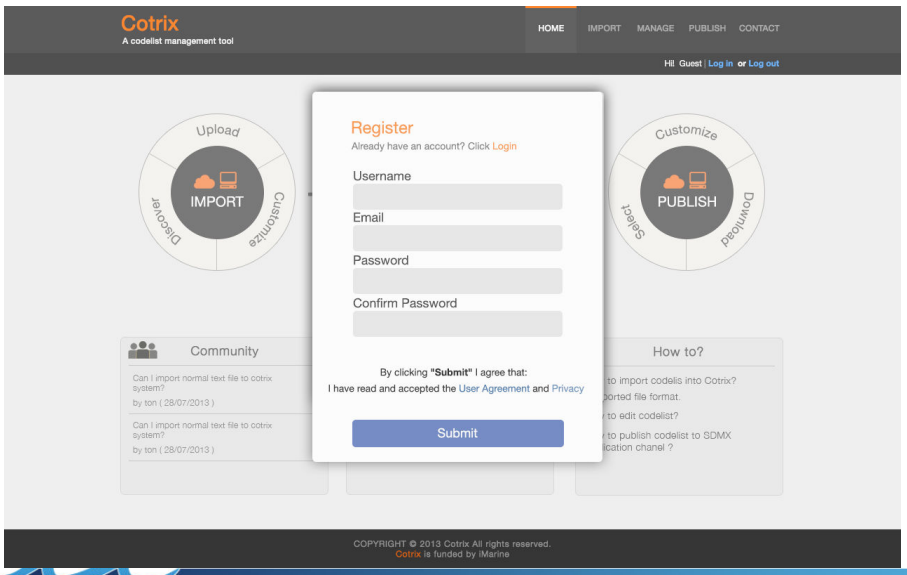


Curate and Inspect the list after the import

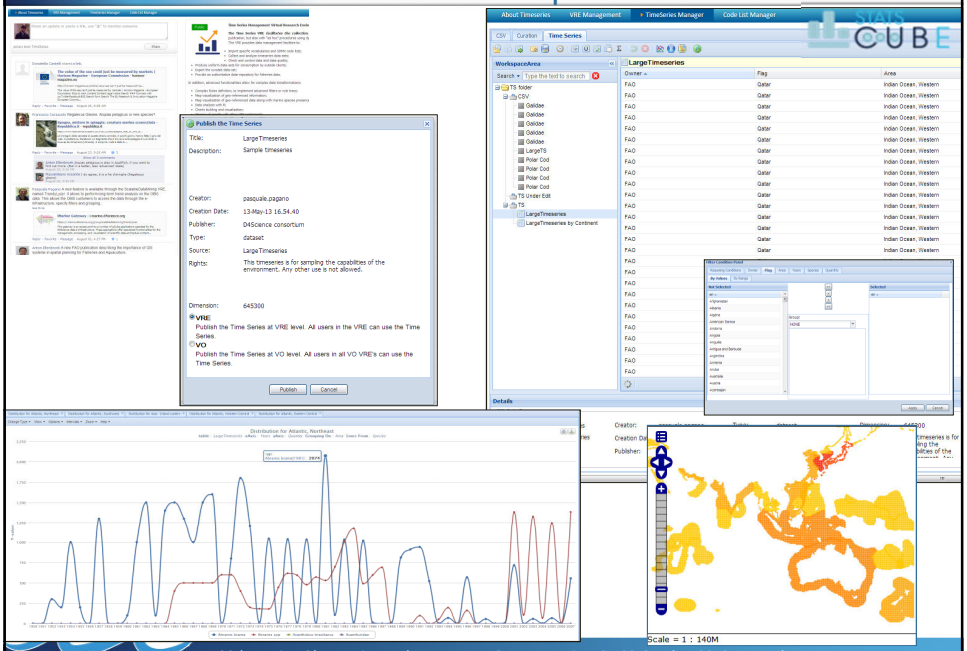
CL_SPECIES

code	en	fr	es	la	AUTHOR
AAA	Adriatic sturgeon	Esturgeon de l'Adriatique	Esturión del Adriático	Acipenser naccarii	Bonaparte 1836
AAB	Twobar seabream	Pagre double bande	Sargo de dos bandas	Acanthopagrus bifasciatus	Forsskål 1775
AAC	Bowfin			Amia calva	Linnaeus 1766
AAD	Yangtze sturgeon			Acipenser dabryanus	Duméril 1869
AAE	Tailjet frogfish			Antennarius analis	(Gosline 1957)
AAF	Lake sturgeon			Acipenser fulvescens	Rafinesque 1817
AAG	Indian mottled eel			Anguilla bengalensis	(Gray 1831)
AAH	Amur sturgeon			Acipenser schrenckii	Brandt 1869
AAI	Chinese sturgeon			Acipenser sinensis	Gray 1835
AAJ	African longfin eel			Anguilla mossambica	(Peters 1852)
AAK	Sakhalin sturgeon			Acipenser mikadoi	Hilgendorf 1892
AAL	Giant mottled eel			Anguilla marmorata	Quoy & Gaimard 1824
AAM	Green sturgeon	Esturgeon vert	Esturión verde	Acipenser medirostris	Ayres 1854
AAN	Fringebarbel sturgeon	Esturgeon barbillons frangés	Esturión barba de flecos	Acipenser nudiventris	Lovetzky 1828
AAO	Atlantic sturgeon			Acipenser oxyrinchus	Mitchill 1815
AAP	Persian sturgeon			Acipenser persicus	Borodin 1897
AAQ	New Zealand longfin eel			Anguilla dieffenbachii	Gray 1842
AAR	Speckled longfin eel			Anguilla reinhardtii	Steindachner 1867

Harmonize: Cotrix



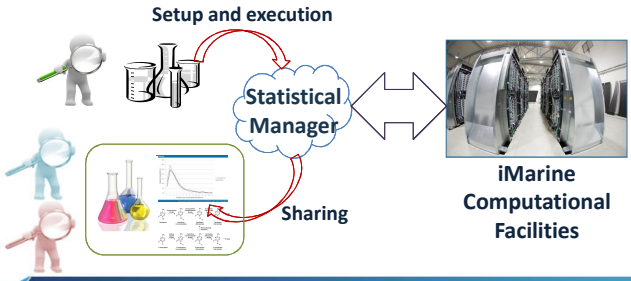
Inspect and curate timeseries



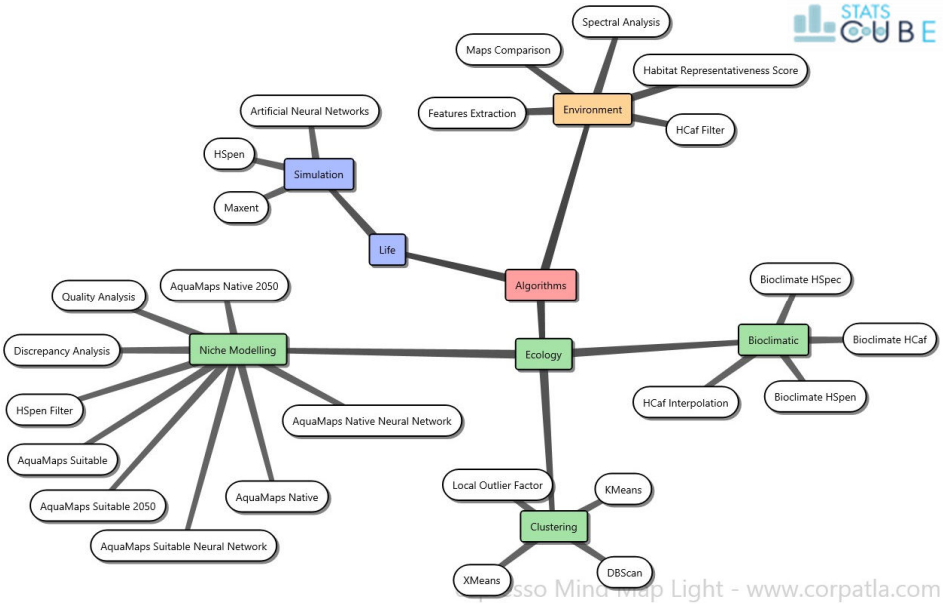
Performing statistical analyses

Facilities helping scientists in:

- Supply ***predefined*** (*reproducible*) state-of-the-art algorithms **as-a-Service**
- Exploit **integrated** *(optimized)* tools such as R, weka, Hadoop
- Perform calculations using e.g. Cloud Computing in a transparent way to the users



Statistical algorithms (21 Oct. 2013)



i-marine

GEOS CUBE

A subset of the products and services belonging to GeosCube

http://wiki.i-marine.eu/index.php/Catalogue:Applications#GeosCube

i-marine

User Services & information products

GEOS CUBE

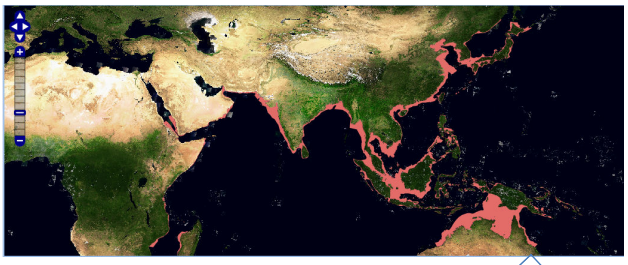
GEOS CUBE

- Visualization of geospatial datasets
- Publication of geospatial information
- Geospatial processes for reallocation, aggregation, interpolation
- Discovery and sharing of geospatial datasets

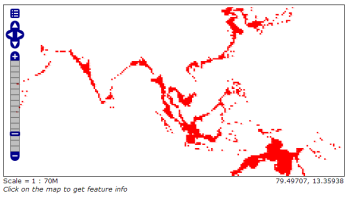
Bridges the gap between powerful infrastructure-based geospatial tools and data, and lightweight web map solutions with limited processing capacity. It thus enables the use of these powerful tools for resource limited users and organizations.

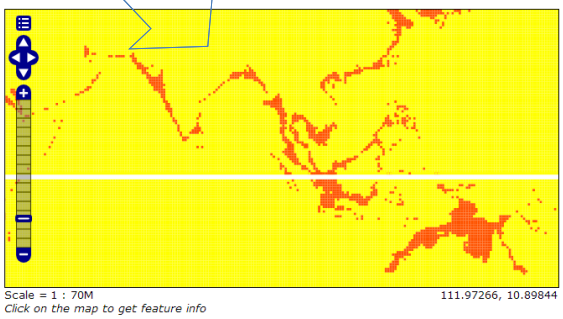
W*Ss, GeoNetwork, GeoServer, THREDDS

Rasterization

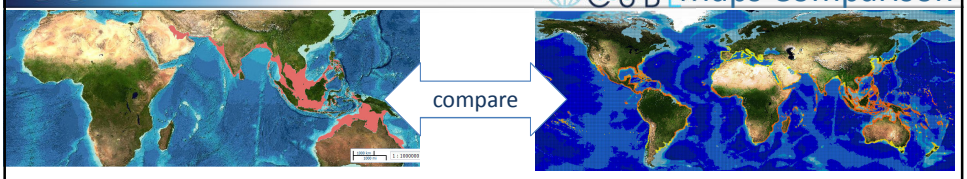


A polygonal map is transformed into a raster map or into a point map



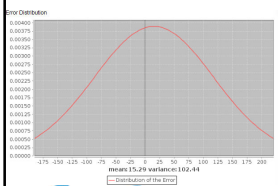


Maps Comparison



Compares :

- Species Distribution maps
- Environmental layers
- SAR Images



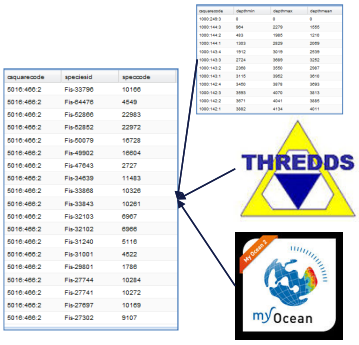
The algorithm produced Multiple Results.

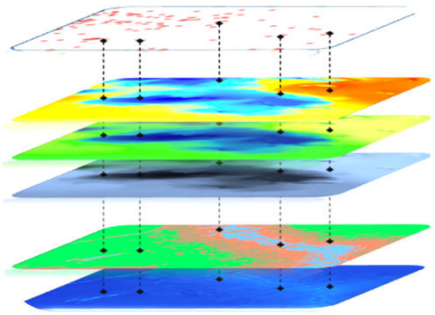
Output Values

MEAN	15.29
VARIANCE	102.44
NUMBER_OF_ERRORS	37512
NUMBER_OF_COMPARISONS	65522
ACCURACY	42.75
MAXIMUM_ERROR	29.95
MAXIMUM_ERROR_POINT	5004:235
COHENS_KAPPA	0.379
COHENS_KAPPA_CLASSIFICATION_LANDIS_KOCH	Fair
COHENS_KAPPA_CLASSIFICATION_FLEISS	Marginal
TREND	EXPANSION
RESOLUTION	0.9972

Environmental enrichment

Environmental enrichment of species occurrence data

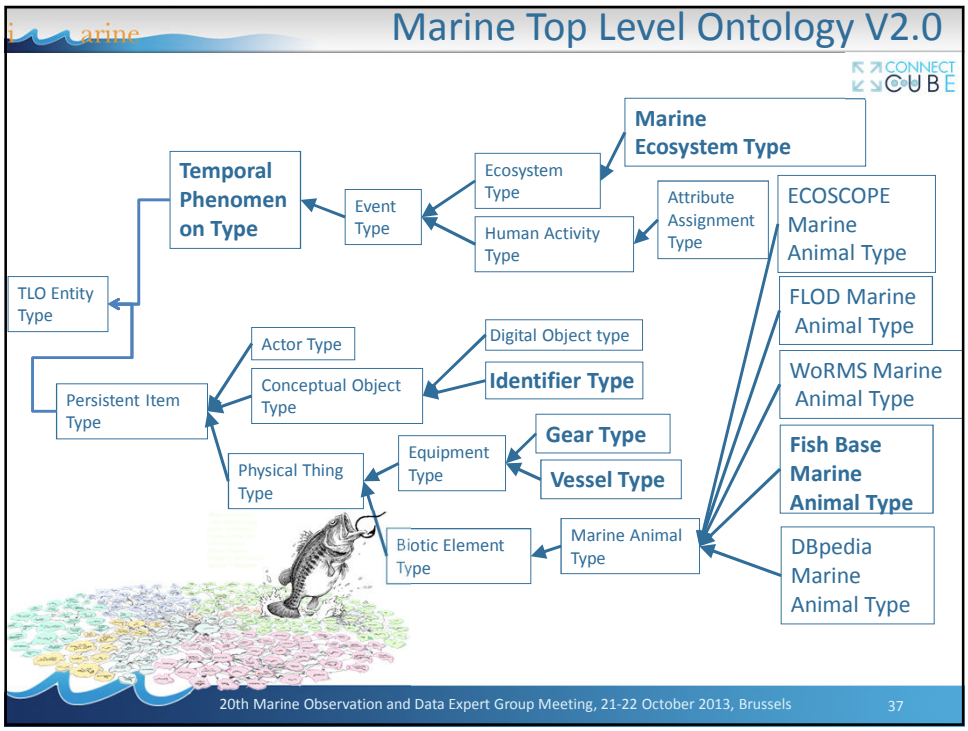


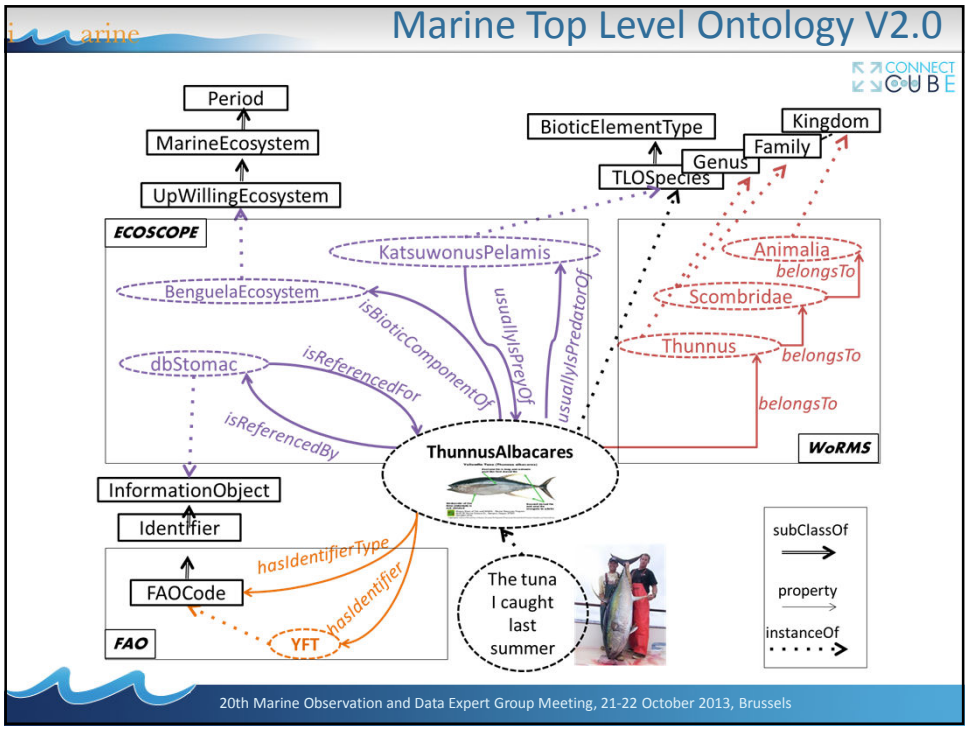


CONNECT CUBE

A subset of the products and services belonging to ConnectCube

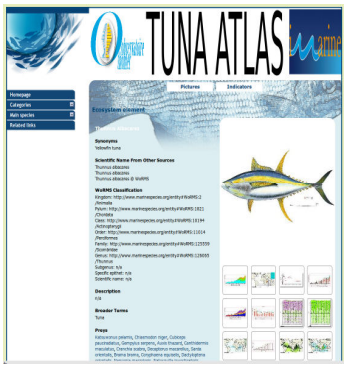
http://wiki.i-marine.eu/index.php/Catalogue:Applications#GeosCube

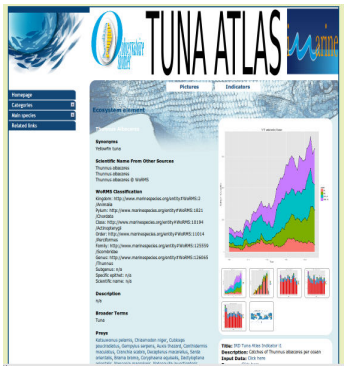

Marine Top Level Ontology V2.0
CONNECT CUBE
Marine Ecosystem Type
Temporal Phenomenon Type
Event Type
Ecosystem Type
Human Activity Type
Attribute Assignment Type
ECOSCOPE Marine Animal Type
FLOD Marine Animal Type
WoRMS Marine Animal Type
Fish Base Marine Animal Type
DBpedia Marine Animal Type
TLO Entity Type
Persistent Item Type
Actor Type
Conceptual Object Type
Digital Object type
Identifier Type
Physical Thing Type
Equipment Type
Gear Type
Vessel Type
Biotic Element Type
Marine Animal Type
20th Marine Observation and Data Expert Group Meeting, 21-22 October 2013, Brussels
37


Marine Top Level Ontology V2.0
CONNECT CUBE
Period
MarineEcosystem
UpWillingEcosystem
BioticElementType
Kingdom
Family
Genus
TLOSpecies
ECOSCOPE
KatsuwonusPelamis
BenguelaEcosystem
dbStomac
isBioticComponentOf
usuallyIsPreyOf
usuallyIsPredatorOf
isReferencedFor
isReferencedBy
Animalia
belongsTo
Scombridae
Thunnus
belongsTo
belongsTo
WoRMS
ThunnusAlbacares
InformationObject
Identifier
hasIdentifierType
hasIdentifier
FAOCode
YFT
FAO
The tuna I caught last summer
subClassOf
property
instanceOf
20th Marine Observation and Data Expert Group Meeting, 21-22 October 2013, Brussels

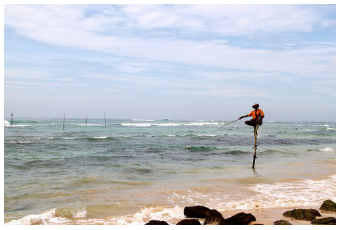
The **Tuna Atlas** use case

Fact sheet: describes an entity (species, vessel..) and its ***characteristics*** (including relationships with other entities) as well as ***linked data*** (images, publications, indicators...) coming from different sources.





Links to additional information (1)



Project Website

www.i-marine.eu

iMarine Gateway

https://i-marine.d4science.org

Board and community pages

http://wiki.i-marine.eu/index.php/IMarine_Liaisons

http://wiki.i-marine.eu/index.php/Ecosystem_Approach_Community_of_Practice:_The_iMarine_Board

Links to additional information (2)

Examples of VREs relying on BiolCube selected services:
- https://i-marine.d4science.org/group/aquamaps
- https://i-marine.d4science.org/group/biodiversityresearchenvironment
- https://i-marine.d4science.org/group/ecologicalmodelling

Examples of VREs relying on ConnectCube selected services:
- https://i-marine.d4science.org/group/documentsworkflow
- https://i-marine.d4science.org/group/fcpps
- https://i-marine.d4science.org/group/isearch

Examples of VREs relying on GeosCube selected services:
- https://i-marine.d4science.org/group/biodiversityresearchenvironment
- https://i-marine.d4science.org/group/ecologicalmodelling
- https://i-marine.d4science.org/group/vesselactivitiesanalyzer

Examples of VREs relying on StatsCube selected services:
- https://i-marine.d4science.org/group/scalabledatamining
- https://i-marine.d4science.org/group/timeseries
- https://i-marine.d4science.org/group/tbti_vre

AppliFish

Species focused information disseminated through mobile applications, for use by anyone at consumer places

Comprehensive knowledge bases **integrating fishery** and **biodiversity** data sources