Open Data in gCube: the iMarine case

Andrea Manieri - Engineering Ing.Inf. Spa
Pasquale Pagano – CNR-ISTI
Anton Ellenbroek – FAO
A journey 10+ years long

The initiative
(the visionary leadership)

The e-infrastructure
(the operational platform)

The system
(the enabling sw system)
Multi-tenant Delivery Model

- Infrastructure as a Service
  - Dynamic deployment
  - Hosting
  - Resource Lifecycle
  - Monitoring
  - Accounting
  - Security

- Software as a Service
  - BiolCube
  - ConnectCube
  - GeosCube
  - StatsCube

- Platform as a Service
  - FeatherWeightStack
  - SmartGears
  - ApplicationSupportLayer
  - SOA3
iMarine exploits a **Hybrid Data Infrastructure** by

- combining *over 500 software components*
- providing access to *more than 25k datasets*
- serving *more than 1000 jobs a day*

**iMarine capacities are offered as services to 1700 researchers in 44 countries**
"Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)" (http://opendefinition.org/)
• **Legal interoperability:** data from two or more databases may be combined or otherwise reused without compromising the legal rights of any of the data sources used.

• **Confidentiality of usage data:** Operation performed by the users are accounted and visible to the VRE and community manager but details are hidden (e.g. Total volume used by the user *but not the file names* or Total number and CPU time used by the user *but not the algorithm used and/or details about the execution*
• **(Digital) Data preservation**: the series of managed activities necessary to ensure continued access to digital materials for as long as necessary. (Source: http://ifdo.org/wordpress/ )
  - Default commitment is for long term maintenance;
  - Criteria of eligibility of standards to establish (by the Community) the format to be supported;
• iMarine Platform commits to:
  - To maintain content through supported metadata;
  - To support a format as long as needed;
  - To support service for a fixed amount of time after decommissioning
  - To notify any service discontinuity
What’s still to be explored?

- **Liability of the infrastructure** for Infringements and violation (ensuring legal interoperability, IPR infringement,
- **Long-term technical** support
  - How to deal with the Increasing amount of storage (specific hardware or sw solutions – e.g. deduplication)
  - How to deal with the Increasing number of formats (complexity in maintenance)
  - How to demonstrate how access rights allows to ensure privacy, confidentiality and security of sensible data
  - How to ensure provenance of data and keep track of their transformation
- **Relevance of data** to be preserved
- **Software maintenance** and its evolution,
- **Costs of the overall infrastructure operation**
All-you-need services

Data

Computing

Applications

iMarine Capacities
Data: Storage as Service

to host and maintain data

**Database**
- High-availability
- Standard
- Ready-to-use

**Cloud Storage**
- Scalable
- Reliable
- Secure

**Geographical DB**
- Scalable
- OGC Standard
- Privacy and Attribution
to curate and manage data

Metadata Generation
- Geospatial Data
- Biodiversity Data
- Statistical Data

Harmonization
- Disambiguate
- Validate
- Integrate and Consistency Check

Data Exchange
- OGC protocols
- DarwinCore
- SDMX
**OAI-PMH, OpenSearch**
- FAO Facksheets
- Aquatic Commons
- Bioline International
- Biodiversity Heritage
- OceanDocs
- Nature, PenSoft Journals
- ...

**SDMX ***
- FAO CodeLists
- IRD CodeLists
- FAO datasets
- Eurostat
- ...

**RDF, OWL**
- FAO FLOD
- Marine Top Level Ontology
- IRD Ecoscope
- FactForge, Yago2
- ...

**DarwinCore / ISO19139**
- >35 M Observations (OBIS)
- ≈ 120 K Observed Species (OBIS)
- ≈ 500 K Taxa (WoRMS)
- >600 K Scientific Names (ITIS)
- >12 K Species Maps (AquaMaps)
- ≈ 600 Species Extent (FAO)
- ... FishBase, SeaLifeBase
- ... CoL, GBIF

**ISO19139 (OGC W*S)**
- >350 variables
- 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, ...

EGI Conference 2015, 21 May 2015, Lisboa
to process and extract knowledge

**Scalable**
Easy to Manage
Across Boundaries
Tailored

**Elastic**
Assignment of Computing
Assignment of Processors
Virtual Research Environment

**Rich and Heterogeneous**
High Throughput
Map-Reduce
Parallel R
Capacities: Computing as Service

iMarine

WPS
Applications as a Service

**AppsCube**
Applications realising a framework conceived to support practitioners willing to develop applications interfacing and benefitting from gCube facilities.

**BiolCube**
Applications to generate new knowledge from data (e.g., modeling and analyzing distribution data, comparing checklists, and producing maps).

**ConnectCube**
Applications offering a comprehensive suite of tools for standards-oriented data publication environment, including semantic technologies.

**GeosCube**
Applications to help practitioners dealing with geospatial information to properly access, consume, and produce data.

**StatsCube**
Applications comprising support for tabular data validation, data enrichment and efficient analytical tools.

**IceCube**
Applications offering core facilities for the deployment, operation and management of a gCube based infrastructure.

A BUNDLE is a set of services and technologies grouped according to a family of related tasks for achieving a common objective.
Bundles used in iMarine

Occurrence and Taxonomic Data Discovery
Occurrence Data Processing
Species Distribution Modeling
Species Distribution Maps Discovery
Taxonomic Data Comparison
Taxonomic Data Matching

Code List Discovery
Code List Management
Statistical Engine
Tabular Data Discovery
Tabular Data Enrichment
Tabular Data Management
Tabular Data Processing

Geospatial Data Discovery
Geospatial Data Processing

Enhanced Documents Management
Fact-sheets Management
Information Object Discovery
Messaging
Shared Workspace
Social Networking Facilities
Virtual Research Environment

to share and collaborate

Share
Database Tables
Workflow
Files

Communicate
Post
Favourite
Connection

Organize
Dynamic VRE Creation
Secure
Policy Control

EGI Conference 2015, 21 May 2015, Lisboa
Methodology

• Common Approach

Import → Harmonization → Generation of Metadata → Publication in Standard Format

• Specialized Implementation

Geospatial Data
Biodiversity Data
Statistical Data
• Import from different sources
• Harmonization and Validation of data
  – spatial and temporal coverage
  – extraction of features
• Generation of metadata
  – Citation
  – Provenance
  – ISO19139
• Publication in Standard Format
  – WFS, WCS, WMS, WPS
• Import from different sources
• Harmonization and Validation of data
  – Status, names,
• Generation of metadata
  – Citation
  – Provenance
  – DwC
• Publication in Standard Format
  – Sharable and accessible through permanent Rest identifiers
- Import from different formats (CSV, SDMX, SDMX files)
- Harmonization and Validation of data
  - spatial and temporal dimensions
  - extraction of features
- Generation of metadata
  - Citation
  - Provenance
  - SDMX
- Publication in Standard Format
  - SDMX*
Take-away elements

• Several communities proven D4Science a suitable platform for their data management

• Any Open Data need to consider also ANY data, to comply with Research Needs

• Multitenant approach, enable by gCube, is key for multidiscilinarity of Science

• Any (open) Science Platform to come, should leverage on gCube legacy
Thanks!

Andrea Manieri, PMP
R&D Business Developer

Engineering
Ingegneria
Informatica spa

Via Riccardo Morandi, 32
00148 Rome
Italy

www.eng.it

Tel. +39-06.83074047
Fax +39-0663074200
Mob. +39-349.2532481
Skype: andrea.manieri
andrea.manieri@eng.it

(source: http://valuesdrivenleadership.blogspot.it/2013/06/new-website-shares-findings-status-of.html)
PRODUCTS AND SERVICES DEVELOPMENT PROGRESS REPORT

A fraction of the products and services belonging to GeosCube
• Rasterization
  – A polygonal map is transformed into a raster map or into a point map

• Maps Comparison
  – Species Distribution maps, Environmental layers, SAR Images

• Periodicity and Seasonality
  – Signal Extraction Tools, Fourier analysis

• Environmental Signal Processing
  – Resampling, Spectogram

• Community-driven
  – SPREAD,
  – Catches per Species indicators: per Ocean / Area, per Fishing Gear type, per Month / Year, and kernel density for biodiversity / ecological datasets (IRD+OBIS+GBIF)
Plot produced by Dr. G.Coro, CNR, Pisa in < 30 mins (starting from a csv)
White shark distribution points; 2 sources

GBIF; consulted dynamically

OBIS; Same species
Different points
Fact-sheet Display

Global inventory of fisheries measures adopted in ABNJ to prevent significant adverse impacts of high-seas bottom fisheries on vulnerable marine ecosystems (VMEs) and associated information.
A community is willing to provide its users with a platform for effectively executing (computational intensive) processes

Goal

A user friendly web GUI
Algorithms (R, Java) can be added
Data provision is straightforward
Steep Learning curve (quick increment of skill)

Strengths

Opportunities

Algorithms automatically exposed in WPS
Large-scale, distributed and flexible computing environment

Weaknesses

Algorithm revision to benefit from computing capacity

Yet another but powerful working environment

Threats

Actions

The community (or its users) should implement the algorithms / processes to offer (minimum requirement)
D4Science.org will then be configured to host and execute the algorithms (Statistical Manager)
The community is willing to expose geospatial data products (including metadata) by maximising potential access and reuse (open science)

**Goal**

**Strengths**
- Opening data via OGC protocols (CSW, WCS, WFS, WMS)
- Generating standard metadata
- A user friendly web-based GUI

**Opportunities**
- Harvesting from CSW services
- Homogenized and fine grained access
- Integrated with other services, e.g. data analytics

**Weaknesses**
- Static data integration

**Threats**
- Either data upload on infrastructure servers
- Or data registration on Infrastructure registry by accepting the Terms of Use

**Actions**
- The community should provide D4Science.org with the data and the related metadata
  - Supported formats (NetCDF, WFS, WCS, Esri-Grid and Geotiff, …)
- D4Science.org will then instantiate and configure a SDI
A fraction of the products and services belonging to BiolCube

PRODUCTS AND SERVICES
DEVELOPMENT PROGRESS REPORT
• **Species Data Discovery**
  – Search across several data providers
  – Search for all occurrences of a set of species and their synonyms
  – Search occurrences for all species belonging a taxon group

• **Occurrence Management**
  – Intersection, Union, Difference, Duplicate Detection

• **Similarity between habitats**
  – Habitat Representativeness Score

• **Community-specific support**
  – Length-Weight Relationships (Time reduction of 95.4%), ...
Preprocessing and Parsing

A flexible workflow approach to taxon name matching

Accounts for:
- Variations in the spelling and interpretation of taxonomic names
- Combination of data from different sources
- Harmonization and reconciliation of Taxa names

Raw Input String. E.g. Gadus morua Lineus 1758

Correct Transcriptions: E.g. Gadus morhua (Linnaeus, 1758)
Define trends for common species
- Account for sampling biases
- Fill some knowledge gaps on marine species

- Most Observed Taxa
- Observation ranks on Large Marine Ecosystems
- Observation ranks on Marine Ecoregions of the World
Trendlyzer – Definition of Common Species

- Trends for common species can be indicators of ecological changes
- A formal definition of common species is not trivial
- A definition based on occurrences distribution gives interesting, result but is affected by sampling biases

<table>
<thead>
<tr>
<th>Mammals</th>
<th>13</th>
<th>15</th>
<th>16</th>
<th>20</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bai phy</td>
<td>Bai phy</td>
<td>Tur tru</td>
<td>Tur tru</td>
<td>Tur tru</td>
</tr>
<tr>
<td>2</td>
<td>Phy mac</td>
<td>Phy mac</td>
<td>Meg nov</td>
<td>Meg nov</td>
<td>Meg nov</td>
</tr>
<tr>
<td>3</td>
<td>Mir leo</td>
<td>Del del</td>
<td>Phy mac</td>
<td>Mir leo</td>
<td>Del del</td>
</tr>
<tr>
<td>4</td>
<td>Bal bor</td>
<td>Bal acu</td>
<td>Gra gri</td>
<td>Del del</td>
<td>Bal acu</td>
</tr>
<tr>
<td>5</td>
<td>Del del</td>
<td>Mag nov</td>
<td>Del del</td>
<td>Del del</td>
<td>Mag nov</td>
</tr>
<tr>
<td>6</td>
<td>Glo mel</td>
<td>Tur tru</td>
<td>Lep wed</td>
<td>Gra gri</td>
<td>Pho pho</td>
</tr>
<tr>
<td>7</td>
<td>Tur tru</td>
<td>Orc ore</td>
<td>Orc ore</td>
<td>Tur phy</td>
<td>Pho pho</td>
</tr>
<tr>
<td>8</td>
<td>Hyp amp</td>
<td>Glo mel</td>
<td>Glo mel</td>
<td>Tur phy</td>
<td>Glo mel</td>
</tr>
<tr>
<td>9</td>
<td>Orc ore</td>
<td>Lep wed</td>
<td>Pho pho</td>
<td>Ste coe</td>
<td>Orc ore</td>
</tr>
<tr>
<td>10</td>
<td>Bal mus</td>
<td>Pho pho</td>
<td>Pho vit</td>
<td>Lep wed</td>
<td>Ste fro</td>
</tr>
<tr>
<td>11</td>
<td>Gra gri</td>
<td>Zai cal</td>
<td>Bal mus</td>
<td>Glo mel</td>
<td>Lob car</td>
</tr>
<tr>
<td>12</td>
<td>Pho pho</td>
<td>Lag acu</td>
<td>Mir leo</td>
<td>Pho dal</td>
<td>Eub gla</td>
</tr>
<tr>
<td>13</td>
<td>Zip cav</td>
<td>Mir leo</td>
<td>Ste coe</td>
<td>Bal mus</td>
<td>Bal mus</td>
</tr>
<tr>
<td>14</td>
<td>Eub gla</td>
<td>Bal mus</td>
<td>Bal bor</td>
<td>Pho vit</td>
<td>Zip cav</td>
</tr>
<tr>
<td>15</td>
<td>Lag alb</td>
<td>Pho vit</td>
<td>Lag acu</td>
<td>Zip cav</td>
<td>Glo mel</td>
</tr>
<tr>
<td>16</td>
<td>Lag acu</td>
<td>Lag alb</td>
<td>Hal gry</td>
<td>Glo mac</td>
<td>Bal bor</td>
</tr>
<tr>
<td>17</td>
<td>Arc pus</td>
<td>Eub gla</td>
<td>Lag acu</td>
<td>Hal gry</td>
<td>Lag acu</td>
</tr>
<tr>
<td>18</td>
<td>Hyd lep</td>
<td>Hyp amp</td>
<td>Lob car</td>
<td>Zai cal</td>
<td>Arc gaz</td>
</tr>
</tbody>
</table>

Grey = not a common species in 1990

Ongoing Activity

EGI Conference 2015, 21 May 2015, Lisboa
A fraction of the products and services belonging to StatsCube

PRODUCTS AND SERVICES DEVELOPMENT PROGRESS REPORT
Complete application for the management of data workflows.

- Data Flow: dataset compliant with a template that is generated and updated in chunks.
- Manage: import, store, transform, validate, access, analyze, visualize, and export.
- Create reports on data activities
A table template defines:
- Table definition
- Columns definition
- A set of harmonization rules*
- A set of validation procedures

Can be applied to any dataset
Can be modified and shared among people

*To be released
### Tabular Data Manager: Panels

#### Task History
- **2014-05-20 14:44:39**: Set Species as a Dimension, referring to species name.
- **2014-05-20 14:44:27**: Set Species = *Cancer*, where *Gen*.
- **2014-05-20 14:39:24**: Set Species = *Cancer*, where *Gen*.

#### Table Data
<table>
<thead>
<tr>
<th>Owner</th>
<th>Flag</th>
<th>Area</th>
<th>Years</th>
<th>Species</th>
<th>Quantity</th>
<th>Expression on ...</th>
<th>Ambiguous values</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>2003</td>
<td>Xiphopenaeus, Tr...</td>
<td>105</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>2004</td>
<td>Xiphopenaeus, Tr...</td>
<td>130</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>2005</td>
<td>Xiphopenaeus, Tr...</td>
<td>228</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>1999</td>
<td>Xiphopenaeus, Tr...</td>
<td>310</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>1958</td>
<td>Xiphopenaeus, Tr...</td>
<td>220</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>1959</td>
<td>Xiphopenaeus, Tr...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>1964</td>
<td>Xiphopenaeus, Tr...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>1956</td>
<td>Xiphopenaeus, Tr...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>1964</td>
<td>Xiphopenaeus, Tr...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>1964</td>
<td>Xiphopenaeus, Tr...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>1964</td>
<td>Xiphopenaeus, Tr...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>1964</td>
<td>Xiphopenaeus, Tr...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>1964</td>
<td>Xiphopenaeus, Tr...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>2001</td>
<td>Xiphopenaeus, Tr...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>1965</td>
<td>Xiphopenaeus, Tr...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>1965</td>
<td>Xiphopenaeus, Tr...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>1965</td>
<td>Xiphopenaeus, Tr...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAO</td>
<td>Costa Rica</td>
<td>Pacific, Eastern</td>
<td>1965</td>
<td>Xiphopenaeus, Tr...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Diagram
- **Sort Ascending**
- **Sort Descending**
- **Columns**
- **Inline Filter**
- **Add Column**
- **Delete Column**
- **Split Column**
- **Merge Column**
- **Change Type**
- **Filter**
- **Replace Batch**
Recipe: take your csv occurrences, select layers from Geonetwork, add your own geotiff
Here: ph and nitrates from World Ocean Atlas
Produce a map plus a statistical analysis in one action
gCube Releases

- **April**: 3.0
  - Templates
  - Validation
  - Transformations

- **June**: 3.2
  - Operations bundle

- **July**: 3.3
  - Operation bundle

- **September**: 3.4
  - GIS
  - SDMX Datasource

- **November**: 3.5
  - Data Analysis
A fraction of the products and services belonging to ConnectCube

PRODUCTS AND SERVICES DEVELOPMENT PROGRESS REPORT
Vulnerable Marine Ecosystems database (VME-DB)

Access the FAO database to update VME fact sheets through the iMarine Reports Manager

EGI Conference 2015, 21 May 2015, Lisboa
The MarineTLO-based warehouse Evolution

Egi Conference 2015, 21 May 2015, Lisboa
- New Version by the end of the project
  - more than 5 million triples
  - providing information for about 50 thousand species
  - data coming from ECOSCOPE, FLOD, WoRMS, DBPedia, FishBase
<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalogue of Life</td>
<td>this data source offers an integrated checklist and a taxonomic hierarchy of more than 1.3 million species of animals, plants, fungi and micro-organisms</td>
</tr>
<tr>
<td>FAO List of Species for Fishery Statistics Purpose (ASFIS)</td>
<td>this includes 12,000+ species of interest or relation to fisheries and aquaculture</td>
</tr>
<tr>
<td>Global Biodiversity Information Facility (GBIF)</td>
<td>this data source offers more than 430 million of records on species and more than 14,000 datasets aggregated from 580+ publishers</td>
</tr>
<tr>
<td>Fishbase</td>
<td>this data source offers access to 32700 Species, 302900 Common names, 53600 Pictures, 49700 References aggregated thanks to the effort of thousand collaborators</td>
</tr>
<tr>
<td>Interim Register of Marine and Nonmarine Genera (IRMNG)</td>
<td>this data source offers access to over 465,000 genus names and 1.6 million species names</td>
</tr>
<tr>
<td>Integrated Taxonomic Information System (ITIS)</td>
<td>this data source offers authoritative taxonomic information on plants, animals, fungi, and microbes of North America and the world</td>
</tr>
</tbody>
</table>

◆ Source cached automatically  ◆ Source hosted  ◆ Source accessed on demand
<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Center of Biotechnology Information (NCBI) Taxonomy</td>
<td>this data source offers a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet</td>
</tr>
<tr>
<td>Ocean Biogeographic Information System (OBIS)</td>
<td>this data source offers more that 37 million records on species and 1,300+ datasets</td>
</tr>
<tr>
<td>SeaLifeBase</td>
<td>this data source offers access to 126000 Species, 27300 Common names, 11900 Pictures, 18200 References aggregated thanks to the effort of hundred collaborators</td>
</tr>
<tr>
<td>World Register of Marine Species (WoRMS)</td>
<td>this data source offers species “names” for more than 200,000 species including 300,000+ species names and synonyms and 400,000+ taxa</td>
</tr>
<tr>
<td>World Register of Deep-Sea Species (WoRDSS)</td>
<td>this data source offers species “names” for deep-sea species based on WoRMS</td>
</tr>
<tr>
<td>SOURCE</td>
<td>DESCRIPTION</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>FAO GeoNetwork</strong></td>
<td>This data source exposes spatial data maintained by FAO and its partners</td>
</tr>
<tr>
<td><strong>World Ocean Atlas</strong></td>
<td>This data source give access to a number of environmental variables. In particular, iMarine focuses on some indicators including Apparent Oxygen Utilisation, Dissolved Oxygen, Nitrate, Oxygen Saturation, Phosphate, Sea Water Salinity, Sea Water Temperature, and Silicate</td>
</tr>
<tr>
<td><strong>Marine Regions</strong></td>
<td>This data source give access to a standard list of marine georeferenced place names and areas including EEZ</td>
</tr>
<tr>
<td><strong>MyOcean</strong></td>
<td>This data source give access to a number of environmental variables. In particular, iMarine focuses on some indicators including ice concentration, ice thickness, ice velocity, mass concentration of chlorophyll in sea water, meridional velocity, mole concentration of dissolved oxygen in sea water, mole concentration of nitrate in sea water, mole concentration of phosphate in sea water, mole concentration of phytoplankton expressed as carbon in sea water, net primary production of carbon, salinity, sea surface height, temperature, zonal velocity, wind speed, and wind stress</td>
</tr>
</tbody>
</table>
### Statistical Data

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IRD UMR EME/Observatoire Thonier</strong></td>
<td>SDMX Registry and Repository: This data source exposes (a) the Sardara database that contains tuna captures data from several countries, aggregated according to CWP statistical squares (1’x1’ or 5’x5’) and (b) the ObServe database that contains tuna and bycatches captures observed by scientific observers onboard French industrial purse seiners.</td>
</tr>
<tr>
<td><strong>SDMX Codelists</strong></td>
<td>SDMX Codelists either directly accessed from the FAO Registry, or manually uploaded through the facility developed in the context of ICIS.</td>
</tr>
</tbody>
</table>

*Source cached*  
*Source accessed on demand*  
*Source hosted*
<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aquatic Commons</td>
<td>offers access to thematic material covering natural marine, estuarine/brackish and fresh water environments</td>
</tr>
<tr>
<td>Biodiversity Heritage Library</td>
<td>offers access to legacy literature of biodiversity held by a consortium of natural history and botanical libraries</td>
</tr>
<tr>
<td>Bioline International</td>
<td>offers access to open access quality research journals published in developing countries</td>
</tr>
<tr>
<td>Central and Eastern European Marine Repository (CEEMar)</td>
<td>offers material covering marine, brackish and fresh water environment</td>
</tr>
<tr>
<td>DataCite</td>
<td>offers access to the same service whose mission is to give access to research data</td>
</tr>
<tr>
<td>DBPedia</td>
<td>contains over 4 millions things including persons, places, creative works, organisations, species and diseases;</td>
</tr>
<tr>
<td>DRS at National Institute of Oceanography</td>
<td>offers institutional publications including journal articles and technical reports</td>
</tr>
<tr>
<td>SOURCE</td>
<td>DESCRIPTION</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Dryad</td>
<td>offers access to the same service whose mission is to give access to research publications</td>
</tr>
<tr>
<td>FactForge</td>
<td>knowledge base resulting from the integration of a number of datasets including DBPedia, WordNet, Geonames, and Freebase</td>
</tr>
<tr>
<td>FAO FishFinder Factsheets</td>
<td>gives access to the Aquatic Species Fact Sheets developed by the same FAO programme</td>
</tr>
<tr>
<td>FAO FLOD</td>
<td>semantic knowledge based hosted in FAO containing a dense network of relationships among the major entities of the fishery domain, including marine species, water areas, land areas, and exclusive economic zones</td>
</tr>
<tr>
<td>iMarine TLO Warehouse</td>
<td>warehouse integrating information from FishBase, WoRMS, ECOSCOPE, FLOD and DBPedia by using the same top-level ontology developed for the marine domain</td>
</tr>
<tr>
<td>Nature</td>
<td>offers access to the articles published by nature.com</td>
</tr>
<tr>
<td>OceanDocs</td>
<td>offers research and publication materials in Marine Science by aggregating content form 256 repositories</td>
</tr>
</tbody>
</table>

- ◆ Source cached automatically
- ◆ Source hosted
- ◆ Source accessed on demand
### Other Data

<table>
<thead>
<tr>
<th><strong>SOURCE</strong></th>
<th><strong>DESCRIPTION</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenAIRE</td>
<td>gives access to the publications aggregated by the same European funded project</td>
</tr>
<tr>
<td>PANGAEA</td>
<td>offers georeferenced data from earth system research via OAI-PMH. The aggregated repositories are 475</td>
</tr>
<tr>
<td>PenSoft Journals</td>
<td>gives access to a number of open-access journals. In particular, iMarine focuses on BioRisk, Comparative Cytogenetics, International Journal of Myriapodology, Journal of Hymenoptera Research, MycoKeys, Nature Conservation, NeoBiota, PhytoKeys, Subterranean Biology, and ZooKeys</td>
</tr>
<tr>
<td>SmartFish Chimaera</td>
<td>knowledge base offering an unified and integrated view on three marine fisheries information sources, i.e. FIRMS – an international knowledge base including fisheries and resource from West Indian Ocean; StatBase – a statistical database containing statistics provided by West Indian Ocean countries; and WIOFish – a regional knowledge base on West Indian Ocean Fisheries.</td>
</tr>
<tr>
<td>WHOAS</td>
<td>offers the production of Woods Hole community including articles and data sets</td>
</tr>
<tr>
<td>YAGO2</td>
<td>knowledge base anchoring entities, facts and events in time and space. The knowledge base contains more than 440 million facts about 9.8 million entities</td>
</tr>
</tbody>
</table>

◆ **Source cached automatically** ◆ **Source hosted** ◆ **Source accessed on demand**