



## "iMarine data platform for collaborations" workshop,

7 March 2014, FAO, Rome, Italy

*Donatella Castelli, ISTI-CNR and iMarine Scientific Director*

*Marc Taconet, Food and Agricultural Organization (FAO), iMarine Board Chair*

*Sara Garavelli and Stephanie Parker, Trust-IT Services Ltd*

### Executive Summary

The sustainable management of fisheries and marine living resources is a societal challenge of global dimensions. No single institution or country can tackle this challenge alone. The European Union's programme for e-infrastructures and data infrastructures has been the pathway to integrating services and tools in a seamless way increasingly at large scale. As a data infrastructure, iMarine ensures computational resources, specific tools and services are open to many different actors and complementary initiatives. The iMarine Board provides a governance framework, which is important for shaping new directions. The Board plays a dual role: 1) it brings expertise from the fisheries, biodiversity and environmental domains, including requirements to shape the tools and services within iMarine and 2) it helps in defining the business cases.

The challenge now lies in defining an effective sustainability plan for the iMarine data infrastructure in the short term and in identifying future opportunities to shape the infrastructure. iMarine has set itself two goals. One goal focuses on defining baseline sustainability to remain operational when funding ends. The other goal centres on a plan for growth.

- **Baseline goal:** ensuring the continued operation of services set-up and usage by the community, which has to fit with this baseline scenario.

#### **Ambitious objective**

Enhancing the services to support stronger policy responses to grand societal challenges, including developing a vision for iMarine as a response to the need for appropriate fishing regulations. Beyond a core community, iMarine has other potential communities with a remit to address societal challenges through an approach that is more user and demand driven and by aggregating the market in different regions of the world.

**Public-centred partnership:** Both the demand and supply side are based on a public-centred partnership with a seed fund required from a core set of public agencies that recognise the need for a global data infrastructure like iMarine.

The iMarine workshop was aimed at showcasing achievements and presenting the concept of a public-centred partnership to support sustainability. To this end, it explored added value for different participants (e.g. users, data providers, and partners) through interactive discussions and surveys.

The workshop was a rich outreach event for the fisheries community, with a format that was particularly effective in obtaining buy-in from the participants. It demonstrated the relevance of iMarine not only to FAO but also other international organisations, highlighting opportunities for new collaborations. **Four user scenarios** demonstrated the **added value of iMarine** in addressing pressing socio-economic challenges related to sustainable fisheries management and the conservation of marine living resources. These user scenarios are described below.

## User scenarios

### 1. Tuna Atlas

iMarine solutions for statistical data managers to harmonise, aggregate and analyse statistical time-series.

**Societal challenges and socio-economic factors:** Tuna and tuna-like species are a significant source of food and income. Their products sustain many livelihoods in both developed and less-developed countries. They include approximately 40 species occurring in most of the globe's tropical and moderate oceans and seas.

**Policy context:** The Tuna Atlas is the annual compilation of global tuna nominal catches, tuna and billfishes catches provided by FAO.

**IT pain point:** This mostly manual process has to deal with heterogeneous data sources, making it particularly challenging.

**Goal:** The goal of the iMarine Tuna Atlas solution is to generate knowledge about tropical tuna, including the generation of indicators.

**Collaborative approach:** Scientists contribute by providing data sources and datasets on fisheries, biodiversity & environment, as well as a set of related processes. iMarine expertise enables service execution, providing servers to combine the data and processes. Virtual research environments (VREs) help improve the data collection process, and enable collaboration between practitioners with different expertise, from researchers in ecology sharing raw processes to a panel of data managers working on methods to improve raw processes.

**Benefits:** Improving the ability to map different data sources and generate Tuna Atlas Fact Sheets in standardised formats, with dynamic execution through a web client. The iMarine data infrastructure ensures a cost-effective approach by making available resources that are not typically affordable by a single organisation. This iMarine solution caters for different computing skills, including data managers using R, a language and environment for statistical computing and graphics.

**Presentations:** [http://i-marine.eu/iME/iME4a\\_-\\_TunaAtlas.pdf](http://i-marine.eu/iME/iME4a_-_TunaAtlas.pdf), and [http://www.i-marine.eu/iME/IRD\\_Tuna\\_Atlas.pdf](http://www.i-marine.eu/iME/IRD_Tuna_Atlas.pdf).

### 2. The Vulnerable Marine Ecosystems (VME-DB) factsheet workflow

The iMarine e-infrastructure as a service provider to support factsheet production including online maps integration.

**Societal challenges and socio-economic factors:** Deep seas play a fundamental role in the marine ecosystem. Understanding the impact of bottom contact gears used by deep-sea fisheries on the sea floor is of utmost importance.

**Policy context:** The United Nations General Assembly adopted resolutions that inspired FAO to co-ordinate the development of the International Guidelines for the Management of Deep-Sea Fisheries on the High Seas.

**IT pain point:** A large amount of dispersed information has been generated by RFMO since 2006 when it developed measures to protect vulnerable marine environments and their coral and sponge communities. It is difficult to access the information that is being entered into a

dedicated database (VME – DB), but this information is important to assist in informed decision-making and the development of further measures to increase sustainability and reduce negative impacts. Overcoming this challenge is also important to facilitate partnerships and create networks among the user community, information contributors and people working on the protection of VMEs in the high seas.

**The goal:** support factsheet production including on-line map integration, ensuring sustainable maintenance. Provide a Content Management System (CMS) to handle VME data in a given data structure, in a specific workflow and within a limited funds available.

**Benefits:** The iMarine Reports Manager provides a flexible CMS for reading content in whatever database schema. It has a user-friendly interface, a role-based workflow and real-time integration. Data protection – the application is instantiated in a dedicated virtual research environment (VRE) for the VME database, which means that access is restricted to selected users as part of the iMarine Terms of Use. Because data is not exposed publicly or shared with other VREs, no specific license is required. Additional community services – the iMarine data infrastructure provides the VME community with a collaborative workspace through a virtual research environment (VRE) to share information and analysis. The user community also benefits from the infrastructure through access to biodiversity, environmental and fisheries data from other VREs.

**Presentation:**

### 3. Chimaera – The Western Indian Ocean Fisheries Regional Portal

This is an example of iMarine solutions for marine practitioners who need to extract information from different data sources and collaboratively produce new products out of this information.

**Societal challenges and social-economic factors:** the requirement to ensure food security and fulfil international commitments to responsible fisheries by improving regional and national fisheries and aquaculture strategies and policies. The SmartFish project aims to provide a portal improve data access and improve knowledge and information used for fisheries management processes.

**IT pain point:** The SmartFish project aims to overcome the difficulty in accessing information and data scattered across independent source information systems.

**Goal:** develop a regional fisheries portal to improve access to dispersed information systems, such as **WIOFish** (a regional knowledge base on West Indian Ocean Fisheries); **FIRMS** (FAO, an international knowledge base including resources from the West Indian Ocean); **StatBase** (a statistical database containing fishery statistics provided by West Indian Ocean countries. Embed iMarine web semantic tools to offer a complete set of complex functionalities with seamless data exploitation and very few data exchanges. Ensure validation through user community usage.

**Benefits:** iMarine acts as a hosting infrastructure that is sustainable and secure. The StatBase hosting solution addresses issues around server instability. The portal embeds iMarine connectCube services in a user-friendly way to guide the user in his/her data search. These services include semantic services for indexing, annotation and search (taking into account each source, data access and sharing policies); creating search indexes and services to exploit the search indexes, and providing query results.

Beyond supporting fisheries managers in their understanding of stocks in a particular area, Chimaera is also suited to data journalists. Features include an interactive map and links to different related sources.

**Opportunity:** This model could be extended to other databases and information systems (regional or national).

**Presentation:** [http://www.i-marine.eu/iME/iME4c\\_-\\_Chimaera.pdf](http://www.i-marine.eu/iME/iME4c_-_Chimaera.pdf).

### 4. BiOnym – a concept mapping workflow for taxon name reconciliation

iMarine solutions for biologists and marine ecologists working with species occurrence and taxonomic data.

**The big data challenge:** Big data is becoming more common in biology, dramatically increasing data volumes. Biologists are already seeing the advantages of good data management, preservation and data sharing, which is transforming the way they do science. Data challenges are particularly difficult in biodiversity because of the need to integrate small, restricted-scope datasets into massive databases through intra-discipline integration (homogeneous) and inter-discipline integration (heterogeneous). The ability to achieve integration is important for tackling societal challenges like climate change by supporting evidence-based management of marine living resources.

**IT pain points:** Taxonomic names are the key to binding together information on the same taxon from different sources. However, different research groups use different spellings; there are human errors (e.g. typos) and errors at the level of the tools used. Reconciliation is a necessity. Moreover, existing systems are not sufficiently flexible, they are often coupled to a single 'reference list' and can lack high throughput needed for large-scale projects.

**Goal:** provide a flexible, highly customisable, workflow-based approach where the user controls input, can use output as input in other processes underpinned by the iMarine data infrastructure.

**iMarine technology success story:** iMarine has successfully integrated COMET, a general purpose concept mapping framework, as part of the BiOnym workflow. Its modular architecture has enabled developers to add new functionalities or to improve existing ones easily.

**Benefits:** A real-world application for concept-mapping principles. Focus on the domain of taxonomy, initially for marine species. It attempts to solve many issues common to the taxonomy community. It is open to integration of third parties components. It is based on standard and open formats and is highly flexible. For example, third-party matchers can be seamlessly 'wrapped' and plugged in the workflow. A near-term goal is to support collaborative matching results evaluation.

**Presentation:** [http://www.i-marine.eu/iME/iME4d\\_BiOnym.pdf](http://www.i-marine.eu/iME/iME4d_BiOnym.pdf)

## The Global Blue Growth Initiative

### FAO Perspectives

The iMarine data infrastructure is aligned with many aspects of FAO's mission. For example, FAO is chartered with providing specific information services, especially to support developing countries. Further, FAO is experiencing increasingly complex demands. There is an important opportunity for iMarine to be the driving force behind these changes. Moreover, iMarine presents an opportunity to scale out across a wider geographical area through FAO members and ensure they have a greater role on the international level.

With regard to future FAO programmes, Indroyono Soesilo - FAO-FI, Director of Resources Use and Conservation Division (FIR), highlighted the added value of the **Global Blue Growth Initiative**.

- **Overview:** FAO is currently developing a Global Blue Growth initiative for the sustainable management and conservation of living aquatic resources in the marine and fresh water ecosystems, as well as in coastal and inland ecosystems, with the aim of supporting food security and nutrition and poverty alleviation.
- **Demand:** There is a high demand for data management solutions to effectively support the sustainable use of aquatic resources and habitats preservation that no organisation in isolation is able to satisfy.
- **Opportunities:** Helping to define and proposing a global partnership/alliance aimed at forging a "Global Blue Growth Data Framework (GBGF)". The iMarine data infrastructure and

its seamless access to diverse types of relevant data, web services, data processing tools, and related expertise is very well suited to supporting such a data framework. The future partnership/alliance could therefore leverage iMarine.

## Horizon 2020 and the Blue Growth Initiative - Call Focus

**Presenter:** Marco Weydert, European Commission

The Blue Growth Focus Area in Horizon 2020 will focus on challenges of marine research and monitoring for a sustainable blue economy.

**Social-economic drivers:** The Blue Growth Economy today employs 5.4 million people but could rise to 7 million by 2020. Blue biotech has an annual growth potential between 5-10%. Offshore wind power generation is expanding rapidly and could meet 4% of electricity demand by 2020. Deep sea minerals extraction could gradually represent up to 10% of the world's minerals and go from virtually zero to EUR 10 billion/year by 2030.

**Focus in H2020:** a new activity 2.5 focused on "cross-cutting marine and maritime research". The aim is to adopt a strategic approach to cross-cutting marine and maritime research to underpin Integrated Maritime Policy and Blue Growth Strategy; catalyse efforts of different Societal Challenges or H2020 priorities to address complex cross-cutting marine and maritime research questions that could not be tackled by a single Societal Challenge and take stock of and advance the FP7 experience "The Ocean of Tomorrow" (31 projects, EUR 195M) on cross-thematic marine and maritime research.

Large and complex cross-cutting initiatives in 5 priority areas: 1) **Sustainable exploitation of the diversity of marine life** 2) **New off-shore challenges** 3) **Exploitation of deep sea resources** – deep sea mining (not covered in 2014-2015) 4) **Technologies and systems for ocean observations**. 5) **Socio-economic aspects**.

**Thematic areas:** ocean energy, aquaculture, biotech, deep sea resources.

**Criteria:** high potential for innovation and growth; genuinely cross-cutting approaches; support across the innovation chain, from research to development; supporting policy (Blue Growth Strategy; International Maritime Strategy; marine framework directive).

## Workshop Takeaways

Presentations and discussions at the workshop highlighted the added value of using the iMarine data infrastructure for research and policy-making.

- **Benefits for researchers.** iMarine has fostered an approach that is driven by user demands, including direct inputs from the iMarine Board. Researchers can benefit significantly from faster time to insight thanks to the iMarine computational power. iMarine has already demonstrated that time reduction goes from weeks to hours. It is also important to note that tools and services (e.g. accessing and sharing data in a unique way) have been built to facilitate the research community without the need to significantly change the way they work though some degree of flexibility is required.
  - Examples of **large computational needs** include **AquaMaps** and **FishBayes**, but are also expected to come from the increasing trend for multidisciplinary approaches (grand global challenges), as well as new virtual research environments and communities through Horizon 2020.
  - iMarine is also a **practical solution** and **platform for open-data compliance**, which benefits not only the users but also data providers by enabling the use of standardised formats and facilitating higher quality data.
- **Advancing the knowledge base.** The iMarine data infrastructure is a powerful technology that fosters support to the ecosystem approach fisheries management and conservation of marine living

resources by significantly advancing the knowledge base necessary to advancing research and informing policy decisions. For example, institutions typically cannot afford to buy computational resources that help accelerate the time to process and analyse data, which in turn lead to new discoveries or insights. iMarine also caters for different programming skills by providing statistical processes in R (increasingly used in biology, as cited by Prof. Carole Goble, Manchester University at the Biodiversity Informatics Horizons 2020, Rome, September 2013) or Java.

- **Connected, collaborative research.** iMarine enables the important shift from privately held data frameworks to collaborative workflows based on the latest state-of-the-art technologies. iMarine supports international collaborative research by connecting researchers with a common goal through Virtual Research Environments (VREs) irrespective of geographical location.
- **Pressing need for standards.** The workshop highlighted the pressing need for global efforts, notably by the fishery community, to elaborate and publish international metadata standards which will ultimately enable interoperability and the ability to efficiently share data on an international level. Standardisation efforts need to focus on harmonisation of semantics, and to consider the entire process and data workflow from the data collection phase (e.g. scientific observers on board fishing vessels), through to regional databases used for joint analysis in support to stock assessments and fishery management, and the production of a dashboard of indicators that support decision making and reporting.

**Standardised formats:** The value of the iMarine data infrastructure comes from improving and harmonising the quality of data from each provider and for each FLUX statistic. This service enables the use of standard formats and run the analytical process based on them.

iMarine provides a comprehensive tool kit for the entire workflow, from raw data to the final analysis made by different practitioners accessing the same environment. Data quality is taken very seriously, adding another important layer to the knowledge base and enabling re-use and further improvements. The iMarine approach has clear advantages not only for data consumers by enabling comparison and contrasting data but also for data providers by improving quality and enabling collaboration between different providers. Examples include monitoring control and surveillance data together with data from a research institute, allowing them both to look at the data and improve it. FAO has also experienced the same benefits, by being able to identify issues and mistakes more easily and making improvements.

- **Cool tools for researchers and citizens.** AppliFish is one example of this because it captures different data sources to increase awareness about marine species and their conservation status from a seafood consumer perspective. The popularity of this mobile app clearly demonstrates its added value in terms of making quality data more widely available.
- **Important opportunities for Training as a Service.** The workshop highlighted the huge potential that exists for training (web-based and f2f) building on the current iMarine offer. Training is seen as one of the incentives for using the data infrastructure and seeing its benefits first hand. This is particularly important in the domain of fisheries, marine biology and environmental sciences, where users and researchers generally lack advanced IT skills. It is important to bear in mind the time to learn to use new tools in that practitioners and researchers come from many different backgrounds. One idea could be to allocate online tutors for clusters of researchers/marine sector professionals with similar needs.