



PROJECT PERIODIC REPORT

Publishable summary

Grant Agreement number: 283644

Project acronym: *iMarine*

Project title: *Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources*

Funding Scheme: CCPCSA

Date of latest version of Annex I against which the assessment will be made:

Periodic report: 1st 2nd 3rd

Period covered: from 1 September 2012 to 30 June 2013

Version: 1.0

Name, title and organisation of the scientific representative of the project's coordinator:

Philippe Rohou, project coordinator, ERCIM

Tel: +33.4.97.15.53.06

Fax: +33 4 92 38 50 11

E-mail: philippe.rohou@ercim.eu

Project website address: <http://www.i-marine.eu>

Deliverable No: D1.3

Submission date: 30 August 2013

Table of contents

<i>1. Publishable summary</i>	<hr/>	<i>5</i>
-------------------------------	-------	----------

1. Publishable summary

IMARINE CONTEXT AND OBJECTIVES

In the marine environment, e-infrastructures play an important role in supporting decision-making processes that have to deal with data from a wide variety of instruments, collection frameworks, and collection needs.

Key needs of marine scientists and policy makers are:

- *Capacity:* users need to access the data, tools and, often, also computational capacity in order to perform analyses or processes.
- *Collaboration:* Multidisciplinary & multi-faceted collaboration at local, national, regional and international levels.

Bringing all the resources required to support research and policy advice requires a governance model that can rely on robust and comprehensive facilities that are also cost effective and accountable. It is important that Europe capitalises on existing data infrastructures while ensuring the long-term maintenance, monitoring and assessment of data.

iMarine is an initiative aimed at supporting the implementation of the Ecosystem Approach (EA) to fisheries management and the conservation of living marine resources. To achieve its objectives, *iMarine* provides a data e-infrastructure that facilitates open access and the sharing of a multitude of data, collaborative analysis, processing and mining processing, as well as the publication and diffusion of newly generated knowledge. This is a complex process because it requires coordination with many actors and initiatives across different scientific and operational domains. It also needs to tackle data heterogeneity while relying on a multitude of resources and technologies, some of which are not yet ripe or powerful enough to meet the given requirements. One of the main challenges that *iMarine* aims to address is data heterogeneity. This means not only tackling data from different disciplines and in different formats, but also dealing with different information and tools that do not belong to any specific community.

THE IMARINE BOARDS & ADVISORY COUNCIL, CENTRAL TO THE INITIATIVE SUCCESS

The *iMarine Board* helps to guide *iMarine's* strategic goals by acting as its governance body and by representing different target communities. These communities come from different domains within the Ecosystem Approach to fisheries management and conservation of living marine resources. Board Members contribute by sharing business cases, helping to define standards, offering advice on the policies that *iMarine* should support and on sustainability issues. To date, among the members of the Board a number of new collaborations have been established, where the availability of the *iMarine* operational data infrastructure has been a fundamental enabler.

The *iMarine Board* members are very active in identifying requirements and validating solutions. In the course of the project the focus of their activity has progressively shifted towards reviewing and commenting the Data Access and Sharing policy document produced by the project partners and on identifying appropriate models for assuring the sustainability of the infrastructure and its governance. The feedback provided by them is being used to orientate management decisions on priority actions required towards these goals and to compile a White Paper on a proposed Public-Private partnership business model to be released during the first part of the third project period.

The Board members also largely contributed to iMarine objectives by liaising with a wider EA-Community of Practice (EA-CoP); i.e. organizations that are not actively contributing to the project, yet stand to gain from the e-Infrastructure facilities that potentially can improve their services, such as by lowering expenditures on software infrastructure reduce the costs of data discovery, harmonization and integration, share development of services, and facilitate the integration of data products in policy processes. More than 20 new collaborations and synergies were established with other projects and initiatives. Some of these collaborations have produced complementary resources, like in the case of the FAO SmartFish and Tuna Atlas projects.

Given the high added value offered to the iMarine activities through the involvement of other stakeholders iMarine has extended the iMarine Board inviting additional experts to join the discussion about the identification of priorities, establishment of synergies and iMarine future directions.

The iMarine Board also helped the design and development of a number of e-Infrastructure resources to be made available to the EA-CoP. These include the Top Level Ontology (TLO), a meta-model for FLOD, ECOSCOPE and WORMS ontologies and facilities for harmonizing complex taxonomic names and for performing trend-analysis across disparate and incomplete data assets.

MANAGING, OPERATING AND ENRICHING A PRODUCTION QUALITY E-INFRASTRUCTURE

In order to meet its objective the data e-Infrastructure must be operated, monitored, and maintained as a 24/7 service based on the policies established by the iMarine Board for the whole duration of the project and beyond. It has to supply services capable of supporting the typical business cases that arise in applying the EA approach. To this aim the project plans to deploy a number of Virtual Research Environments (VREs) offering advanced facilities. These range from statistical management of socio-economic data, including data on aquatic ecosystems and monitoring of vessels, fleets, and activities, to management, transformation, and visualization of different types of marine biodiversity data including species data and taxonomies. CoP members will be enabled to exploit these facilities on-demand either programmatically or through innovative applications. The e-Infrastructure will provide these services by leveraging services provided by other existing e-Infrastructures.

The infrastructure was upgraded five times in order to deploy new components and enhancements and site managers were involved in the installation of new nodes for the production infrastructure in order to run both gCube and third-party services. Several other components were also installed and configured to empower the e-Infrastructure for dealing with a richer set of resources, like a Cassandra cluster, a Virtuoso repository, a Thredds server and a number of CouchBase servers. Resources dedicated to the Hadoop and MongoDB clusters were extended and important databases, like FishBase and SeaLifeBase, were installed.

Monitoring tools based on the de-facto standards Nagios and Nagios-NRPE were deployed and configured to operate not only on the gCube component but also on the external ones

integrated in the infrastructure.

VRES, NEW RESOURCES AND TOOLS TO SERVE THE COMMUNITY

The number of users of VREs continues to grow. On average, it has been observed an increase of +34% (147 at Jul '12 vs 197 at Jun '13) in the number of users served by FARM VREs and of +114% (112 at Jul '12 vs 240 at Jun '13) in the number of users served by gCubeApps VREs.

Work on the Virtual Research Environments has led to the operation of 14 cooperative environments aiming at serving various scenarios and, overall, more than 460 users. In particular six new VREs, listed below, were created in the period. These, and all others, are accessible through the iMarine gateway (<https://portal.i-marine.d4science.org/web/guest>)

- *FishFinderVRE*

The explicit purpose of this VRE is to enable 50 authors to prepare hundreds of species fact sheets by offering facilities to elaborate Species Fact Sheets, fill / view their metadata, and select data for download and-or display in a Stand-alone version of the VRE. The tool is operated by one VRE manager in FAO, communicating with dozens of globally distributed authors.

- *iMarineBoard*

Designed to provide the members of the iMarine Board with collaboration tools and a demonstration of infrastructure facilities. In particular, it includes: (i) a collaboration suite including a shared workspace and messaging system, (ii) services for accessing biodiversity data from several major databases, and (iii) services for managing tabular data (e.g. catch statistics) and code lists.

- *VME-DB*

Serves fisheries and aquaculture authors willing to collaboratively produce Fact Sheets on Vulnerable Marine Ecosystems (VME). The VRE supports the definition of new report templates and production workflows as well as the collaborative creation of new reports compliant with defined templates.

- *BiodiversityResearchEnvironment*

Designed to provide its users with a number of facilities to access and manage Biodiversity data, it is equipped with a service to discover species data (including occurrence points) from various data providers including Catalogue of Life, GBIF, ITIS, IRMNG, OBIS, WoRMS, and WoRDSS.

- *ScalableDataMining*

Supporting the application of Data Mining techniques to biological data, it enables users to first select their data using a template for pre-filtering and then apply an identified algorithm to the selected data. It also offers the resources required to perform the computation in a distributed fashion on the e-Infrastructure nodes or on local multi-core machines, and a process monitoring that allows the user to track the process, intervene when necessary, and manage the output. This three-phased approach to computational requirements is extremely versatile, and can be extended by users wishing to bring their own data and algorithms.

- *TCom*

Deployed to serve project members forming the Technical Committee. It aims at facilitating a more efficient and collaborative use of both knowledge based and technology based resources to the benefit of the community of scientists, managers and policy advisors working in the Marine Ecosystem management domain.

MORE AND BETTER SERVICES TO THE COMMUNITY

The extent to which a data e-Infrastructure succeeds as an application platform for a Community of Practice is largely dependent on the tools and services that it makes available to develop and execute the broad class of applications for which it is intended. For the e-infrastructure supporting the iMarine community the target applications are those required to implement the EA, and thus to execute, collate, access, integrate, annotate, transform, search, curate, and publish a variety of data types, including statistics, biological records, environmental observations, and semantic structures. This functionally rich e-Infrastructure will be obtained by integrating and enriching a large number of technologies across the services and tools of gCube, an open software system developed in the framework of other EU funded projects.

The following major technical achievements can be highlighted:

- The *Tree Manager gCube Service* enriched to meet new requirements to support a larger variety of datasets and documents representations.
- The *Species Product Discovery Service* enhanced with new facilities and connectors to several biodiversity data providers. A query language for interacting with this service in easy way has also been designed.
- *Data Transfer* empowered with a portlet that is fully interactive with the iMarine Distributed Storage facilities allowing users to browse the storage and transfer files from one infrastructure node to another or from remotely hosted files to the infrastructure nodes. Computational facilities have also been released, based on the Hadoop technology, which are able to access to remote datasets and to transfer and process them through a map-reduce approach.
- *Quality Assessment and Harmonization of data* improved mainly in the facilities related to SDMX and Tabular Data System. SDMX data can be now exchanged, SDMX messages can be produced and processed, and SMDX time series datasets can be harmonized with respect to reference datasets (e.g. codelists). On the other side the Tabular Data System that has been until now designed addresses the management of time series with a novel approach aiming at enhancing efficiency and data representation effectiveness.
- *Information Retrieval* improved towards providing a distributed and scalable solution, based on emerging technologies and provision of uniform results ranking in the most complex and challenging scenarios available in the infrastructure.
- *Data Manipulation* enriched with capabilities for processing new data types and supporting the new content management layer. The gCube Data Transformation Service is now integrated with the Tree Manager, personalization related services use the Storage System and a pluggable service of the GIS publisher allows the publishing of GIS data into Geoserver/Geonetwork/Thredds.
- *Statistical Manager* redesigned in order to deal with more complex inputs and outputs adopting a distributed architecture for the incoming requests. Several enhancements on TimeSeries Analysis and Management have been implemented that provide support, among the others, to environmental trends calculation, analysis of catch statistics in order to detect periodic phenomena, new methods for taxa matching, and maps comparison. Effort has been also placed on OGC maps processing, while several algorithms for general purpose modelling have also been introduced. The new redesign includes also Trendlyzer, a tool supporting the finding of biodiversity indicators from OBIS.
- *Semantic Data Analysis* improved in terms of architecture, integration, scalability and openness. The latest version of the XSearch Service is now able to exploit the SPARQL

endpoints of the MarineTLO (Top Level Ontology) warehouse. The design of a new tool, XLink, able to analyse the contents of a document and enrich its entities with semantic information based on a SPARQL accessible knowledge base has also been initiated.

DIGITALLY REACHING OUT WITH APPS, TRAINING AND WEB-BASED MEDIA

iMarine is largely based on a cross-disciplinary and community-centric approach. In order to foster interest and cohesion between the members of the CoP appropriate communication, dissemination and training activities have been put in place and are updated frequently, in line with iMarine developments. These activities aim at leveraging and expanding on partner networks by forging strategic alliances in Europe and globally.

Further to the wide and frequent communication of iMarine achievements and developments via the web (www.i-marine.eu) platform, event participation & organisation, news items and interviews. Two activities in this period merit highlights:

- *iMarine mobile application “AppliFish”*: On January 2013 the first iMarine mobile application has been published both for Android <https://play.google.com/store/search?q=applifish> and iOS <https://itunes.apple.com/ca/app/applifish/id593857305?mt=8&ign-mpt=uo%3D2>. The application provides over 550 informative fact sheets on marine species combining data from different global and authoritative data providers (FAO-FishFinder, WoRMS, Fishbase, SeaLifeBase, IUCN, AquaMaps, OBIS). AppliFish has been developed as a demonstrator of the iMarine platform capacity and it has been exploited as a ready for use product to raise awareness on the projects objectives towards policy makers and to create awareness towards the general public.
- The *e-training* structure has been defined and organised in four following courses (plus a section dedicated to webinars): “About iMarine”; “Creating a collaboration portal”; “e-Infrastructure management facilities”; “iMarine applications”. Each training module focuses on the iMarine infrastructure and applications providing a set of user-friendly videos with step-by-step and clear explanations. Each module aims to inform and educate specific target users: Data managers, Scientists, Data Infrastructure managers.

WHO IS BEHIND IMARINE?

iMarine management is jointly carried out by two individuals:

- *Philippe Rohou* – Administrative and Financial Director, and contact point to the European Commission
- *Donatella Castelli* - Project Director & Scientific / Technical Coordinator

More information about the project can be found on the project Website: <http://www.i-marine.eu>

Thirteen contractors participate in the project activities:

