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*: to access 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 relying 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 sharing of a multitude of data, collaborative analysis, processing and mining processing, as well as publication and diffusion of newly generated knowledge. This is a complex process because it requires coordination with many actors and initiatives across 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. In order to meet these requirements *iMarine* has developed novel technologies offering harmonization and cross-domain analysis. The outstanding data processing features implemented in *iMarine* now enable the Ecosystem Approach Community of Practice (EA-CoP) to have rich information sets for multi-dimensional analysis in a collaborative and shared environment rather than in isolated silos. Differently from other e-infrastructures, *iMarine* focuses on the entire data-processing workflow, and stores data access settings, process, and results in the same infrastructure. This is always done in agreement with established data policies.

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

The iMarine Board assists in defining *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.

In the course of the project iMarine Board Members have contributed by sharing business cases, helping to define standards, validating implemented solutions, offering advice on the policies that *iMarine* should support and suggesting mechanisms for its sustainability.

An Extended iMarine Board was formed at the end of the second project period including 29 new experts with the role of "Observers". The members of this extended board have been very

active in helping establishing new liaisons. With a multiplicative effect this support has been growing over time.

As a result of support by the Board along with the many face-to-face meetings held, nine major collaborations and synergies have been established with other projects and initiatives. Some of these collaborations have concerned the co-development and co-funding of tools, e.g. CONTRIX, VM-DB, SmartFish.

The contribution of the iMarine Board was also central to the shaping and validating of integrated solutions for data harmonization. These include, among other, the Top Level Ontology (TLO) that was published during the final project year in several public ontology and vocabulary registries in order to make it also available to others. The TLO semantically extends the underlying models of existing marine data sources including FLOD, ECOSCOPE, WoRMS, FishBase and DBpedia. During this period another element contributing to data harmonization solution was finalised, i.e., a powerful code list management tool capable of supporting the entire life-cycle of code lists from their creation to exploitation. Support to standards for statistical data (SDMX), biodiversity data (Darwin Core), and geospatial data (OGC) has been improved. Notably, OGS WPS standards has been adopted for publishing processing facilities under standard description.

The iMarine Board members have largely contributed to reviewing and enhancing the Data Access and Sharing policy document produced by the *iMarine* project and, especially during the last year of the project, on identifying appropriate models and markets for assuring sustainability and governance of the infrastructure. The feedback provided by the iMarine Board has been used in preparing the Business and Sustainability Plan. This document contains a detailed analysis of costs and possible sustainability models for the *iMarine* infrastructure. These are studied with respect to the requirements and offers of the *iMarine* stakeholder organisations, and includes a baseline model as well as an identified future growth scenario.

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* initiative. It has to supply services capable of supporting the typical business cases that arise in applying the Ecosystem Approach. To further contribute to this aim the deployment of additional Virtual Research Environments (VREs) offering facilities meeting the requirements of newly identified application scenarios has been performed together with the maintenance of those put in operation in the previous project periods. These VREs range from supporting the execution and repetition of experiments to providing working environments supporting the management of tabular data as well as environments for retrieval and semantic discovering of information objects from heterogeneous data sources.

Community of Practice members are enabled to exploit these facilities on-demand either programmatically or through the supported virtual collaboration environment. At the end of the reporting period VRE users were approximately 770.

In the last project period the infrastructure was upgraded five times in order to deploy new components and enhancements, while site managers were involved in the installation of new nodes for supporting specific services and tools (e.g. Marine TLO warehouse, R and Rstudio, SmartFish, OBIS DB node).

The extent to which a data e-infrastructure succeeds in providing a service of good quality in a sustainable way largely depends on the infrastructure enabling services responsible for managing and orchestrating all the other ones. Several activities were performed in the reporting

period for improving these services, making them more flexible and less demanding when being integrated with third-party solutions.

First of all infrastructure resource monitoring and accounting tools were empowered by both developing a new set of facilities and exploiting solutions available on the market (e.g Ganglia).

Important components were also developed, examples of which are SmartGears, a set of libraries which can transparently turn servlet based containers and web applications into gCube resources; a module supporting Distinguished Name based Authentication and another one for SAML Identity Federation; a Workflow Engine enhanced with HiveQL and an Orchestrator Layer. The Search service was extended to operate also over the user's workspace files while the X-Search service was enriched with a library, X-Link, supporting updating of the categories of entities, inferring the connectivity of the detected entities, and enriching them with semantic information.

VRES, NEW RESOURCES AND TOOLS TO SERVE THE COMMUNITY

As of September 2014, 20 VREs were active. Six of them are new VREs added in the third project period to support specific collaborative activities. These are accessible through the *iMarine* gateway (https://portal.i-marine.d4science.org/web/guest).

- *BiodiversityLab*: supporting the needs of scholars willing to perform experiments (e.g. species distribution maps production, species data inspection) on single individuals or groups of marine species;

- *BiOnym*: providing scientists dealing with species names with a service (ByOnym) to compare a set of scientific names against taxonomic reference lists including recognised ones;

- *iSearch*: offering information retrieval and semantic web facilities for seamlessly discovering of information objects from heterogeneous data sources;

- *MarineSearch*: providing information retrieval and semantic web facilities for seamlessly discovering of information objects from marine-oriented heterogeneous data sources;

- *TBTI*: serving the members of the *Too Big to Ignore* initiative with an environment for experiencing with the gCube services;

- *TabularDataLab*: implementing a working environment supporting the management of tabular data, including facilities for collaboratively perform activities like import, curation, analysis, visualization and publishing of tabular data resources.

All VREs deployed in the previous period have been substantially enriched with new generic or empowered services. In particular, a number of specific processes and algorithms have been made available by integrating them as part of the statistical service offering and by exposing them via WPS processes.

In the last part of the project, VREs addressing the needs of new external user communities were also developed:

- *SmartFish*: focused on building a new portal giving seamless access to information and data scattered across three data sources (WIOFish, FIRMS, and StatBase) http://smartfish.d4science.org/.

- *EGIP*: aimed at building an access point to data and information related to geothermal energy. https://egip.d4science.org.

- *Social-ISTI*: providing the members of the ISTI research institute with an innovative working environment. https://social.isti.cnr.it.

The number of users has been growing continuously. On average, an increase of 41% in the number of users served by FARM VREs has been observed (197 in June 2013 vs 278 in September 2014) and of 105% in the number of users served by gCubeApps VREs (240 in June 2013 vs 492 in September 2014).

In order to satisfy the functional needs of these VREs, a set of new software artefacts were released or substantially improved. The new ones include *BiOnym*, a taxonomic name matching application; *COTRIX*, a collaborative code lists management system; *Tabular Data Manager*, a comprehensive environment for tabular data management; and *TrendyLyzer,* an environment for discovering and visualising trends in species databases. Among the improved software artefacts are the set of *Social Networking Facilities* and the *Shared Workspace.*

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 Community of Practice, appropriate communication, dissemination and training activities have been put in place which are being frequently updated, 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, a considerable number of diversified activities were performed: press&media (25), newletters (13), posters (5), press-releases (3), presentations shared on SlideShare (12 presentations, 2000 views), scientific papers (20), interviews (2), participation in external events (30), invitation-only scientific workshop (3).

A new version of the *iMarine mobile application "AppliFish"* was released on January 2014. This app has had 8400 overall downloads in total (6335 for iOS and 2124 for Android), and the number has doubled since the last reporting period. The web version of AppliFish[1] is one of the official information sources of the "Inseparable" portal of the EC[2].

The e-training activities include: (i) organisation of two courses at the UPMC - Sorbonne Université in Paris, and at the Observatoire Océanologique de Villefranche sur mer, as part of the Master in Océanographie Environnements Marins (OEM) (66 students completing the courses); and (ii) webinar hosted by two of the premier sources of information about coastal and marine planning and management tools in the United States and internationally, namely the OpenChannels.org and the EBM Tools Network (over 90 researchers, ecologists, marine biologists, students and analysts from 23 different countries); and (iii) 3rd MUMIA Training School on Information Retrieval (54 students from 17 different countries).

WHO IS BEHIND IMARINE

iMarine is driven by a consortium of research centres, international organisations and small and large enterprises. The scientific and technical management is led by Dr. Donatella Castelli, Project Director. For more information, please contact info@imarine.eu.

---

[1] http://www.i-marine.eu/AppliFish/Default.aspx

[2] http://ec.europa.eu/fisheries/inseparable/en/find