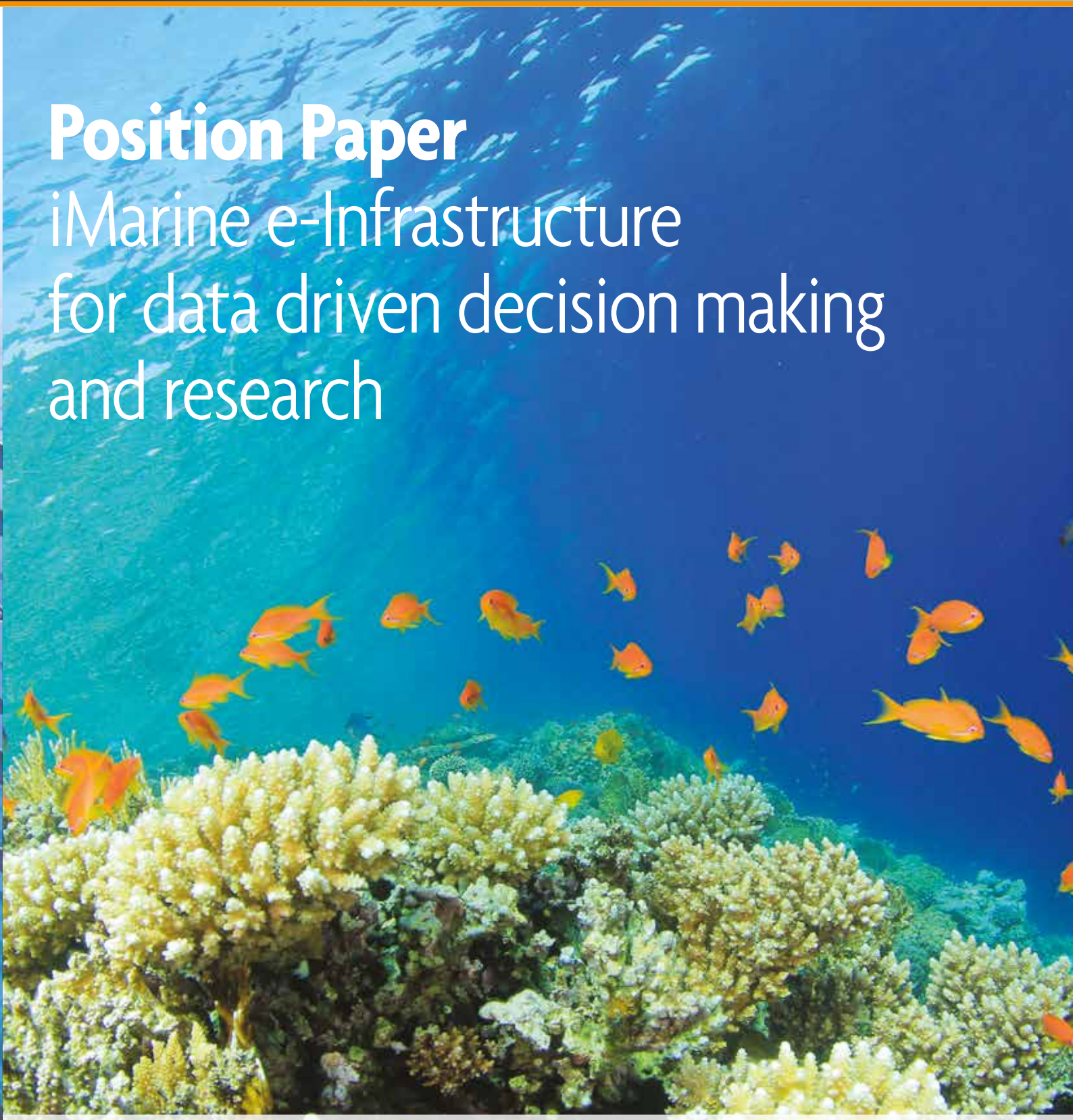
Position Paper

iMarine e-Infrastructure for data driven decision making and research



NOVEMBER 2013

In the marine environment, data 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.

The research leading to these results has received funding from the European Union Seventh Framework Programme ([FP7/2007-2013]) under grant agreement n° [283644].

About iMarine

www.i-marine.eu

iMarine is an initiative aimed at supporting the implementation of the Ecosystem Approach to fisheries management and the conservation of living marine resources.
iMarine is funded by the European Commission, DG-Connect unit, under Framework Programme 7. It involves thirteen international partners:

» GEIE ERCIM, France (iMarine Coordinator)
» Consiglio Nazionale delle Ricerche (CNR), Italy (iMarine Scientific and Technical Coordinator)
» National and Kapodistrian University of Athens (NKUA), Greece
» European Organisation for Nuclear Research (CERN), Switzerland
» Engineering Ingegneria Informatica SpA (E-IIS), Italy
» Foundation for Research and Technology Hellas (FORTH), Greece
» Terradue s.r.l. (Terradue), Italy
» Trust-IT Services Ltd. (Trust-IT), United Kingdom
» The Food and Agriculture Organisation of the United Nations (FAO), Italy
» Fishbase Information & Research Group Inc. (FIN), Philippines
» United Nations Educational, Scientific and Cultural Organization – UNESCO (UNESCO), France
» Centro de Referencia em Informacao Ambiental (CRIA), Brazil
» Institut de Recherche pour le Developpement (IRD), France

iMarine provides an 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 dissemination 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 is also important 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.

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 (Communities of Practice). 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, Board Members have been instrumental in establishing several new collaborations, where the availability of the iMarine operational data infrastructure is a fundamental enabler.

The ecosystem challenge

Aquatic ecosystems, inland, coastal and marine, provide humans with resources for food, livelihood and recreation. They also perform important environmental functions that contribute to general human well-being.

Achieving sustainable use of aquatic ecosystems has been the main and largely failed objective of fisheries management for decades.

The ecosystem approach to fisheries, or EAF for short, imposes further efforts in this direction, implying better understanding and better governance.

The enabling role of data infrastructures

Scientific research depends on tools, instruments and socio-economic systems that help organise and share knowledge. Information and Communication Technologies (ICT) enable close collaboration in almost real time between scientists all over the world, providing access to unprecedented volumes of scientific information processed on powerful computing platforms. Today, ICT-based infrastructures (e-infrastructures) have become a fundamental foundation of research and innovation.

Europe has made significant investments in the development and deployment of e-infrastructure as a key enabling technology for scientific discoveries and to help tackle 21st-century grand societal challenges. Open data and open science are essential to boost innovation and demonstrate the return on investment by national governments and the European Union to society at large. With more investments to incentivise openness and reward contributors, Europe and other world regions will be able to take huge steps forward in making research data more reproducible, gaining transparency and easing collaboration on existing research data.

In the marine environment, data 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. Bringing all these resources together into a policy framework requires a governance model that can rely on robust and comprehensive facilities that are also cost effective and accountable.

iMarine Vision

Richer, better quality and timely science-based knowledge related to aquatic resources will be available through the iMarine data e-infrastructure. The aim is to serve policy development and implementation towards goals of the Ecosystem Approach. The data e-infrastructure will enable a cost-effective and facilitated retrieval, access, collaborative production and sharing of information and tools.

By interconnecting all concerned actors operating in different domains around common data, information and multidisciplinary knowledge-building, iMarine will facilitate the emergence of a more unified and effective Ecosystem Approach Community of Practice (EA-CoP).

iMarine: a powerful data infrastructure for marine ecosystems

Today, 88% of fish stocks in the European Union are fished beyond their maximum sustainable yield, which has led to a steady decline in average fish size over the past 20 years. Achieving good environmental status by 2020 is a requirement of the European Marine Strategy Framework Directive. Avoiding significant adverse impacts on stocks, species, and ecosystems is crucial to this goal.

Data challenges

Marine data is essential to understanding marine ecosystems. Yet obtaining data is much more challenging than it may seem. Data is usually collected for different purposes and often varies in quality. Integration and processing might be computationally intensive and time-consuming, which no single researcher can afford to do on his or her own. Overcoming these barriers to discovery is central to facilitating science as an open enterprise.

Data infrastructures also need to be interoperable while guaranteeing service usability, which is challenging because data is spread across different providers. Another challenge is data location and ownership: how do we ensure that data is openly accessible and fit for use? Interoperable and standardized data products and services ensure that data is available in harmonised formats."

A truly global initiative

iMarine has developed a powerful data infrastructure to address the challenges of fisheries management and the conservation of living marine resources imposed by the ecosystem approach. A key goal of iMarine aims to support stakeholders in the marine ecosystem, spanning biologists, data managers, e-infrastructure specialists and now also citizens through its mobile app, AppliFish.

iMarine is a truly global initiative, not only through its partnership, pool of experts and synergies but also in terms of its data providers. iMarine biodiversity data providers include the Catalogue of Life; the Global Biodiversity Information Facility; the Interim Register of Marine and Non-marine Genera: the Integrated Taxonomic Information System; the National Centre for Biotechnology Information; Ocean Biogeographic Information System and the World Register of Marine Species. iMarine environmental and climiate change data providers include Earth Science Data and Processing Commodities; MyOcean and WorldClim. iMarine socio-economic data providers are the Food and Agriculture Organization (FAO) and GeoNetwork.

Delivering economies of scale – reducing investment costs

Developing e-infrastructure for scientific data is sometimes a costly process but there is a significant element of public good in this, especially when e-infrastructure help tackle important societal and global challenges.

One of the main objectives of iMarine is to reduce costs by capitalising on earlier investments in e-infrastructure through the integration of existing solutions as key enablers. iMarine builds on the legacy of the D4Science data infrastructure for management and operation (e.g. 24/7 operation and monitoring). It also continues to develop the gCube software, a framework for building data infrastructures that can support different scientific commu-

nities. A key enabling factor of gCube is the ability to achieve economies of scale by sharing services or by making the same service instance available to multiple communities, reducing both deployment and operational costs.

By transforming data and returning it in a standard format, iMarine is able to provide additional value-add services for the validation of metadata. This approach can help national fisheries organisations gain economies of scale.

Fostering open access and data sharing

Open data and open science are essential to boost innovation and demonstrate the return on investment by national governments to society at large. The iMarine data infrastructure is designed to facilitate open access, the sharing of data, collaborative analysis, processing and mining processing, as well as the dissemination of newly generated knowledge.

Science as a Service

In the marine environment, scientific research plays a key role in improving our knowledge of aquatic ecosystems in the drive towards efficient fisheries management and sustainable use. Much of this highly challenging research takes place collaboratively. By facilitating data discovery, exchange and harmonization for scientists and practitioners, iMarine enables the exploitation of massive data volumes and computing capacity to support this goal. This approach is based on the concept of open science, to enable new ways to share, compare, and discover data.

Creating innovative research environments that offer everything a scientist needs 'as-a-service' is a game-changing process. The idea is to enable scientists to focus on their core work, without needing to know about the underlying technologies. Researchers only see 'off-the-shelf' facilities for data access and processing catered to their specific application domain.

Compelling catalogue of applications

BIOL CUBE GEOS CUBE CONNECT CUBE STATS CUBE

The iMarine infrastructure is like an archipelago where applications emerge as islands of services resting on an underlying infrastructure bedrock. While each island focuses on a specific application domain, the archipelago is an interconnected system, enabling marine researchers to go 'island-hopping' whenever they need a purpose-built application for their work.

The platform can host a compelling portfolio of applications that help generate new knowledge of the marine ecosystem. The current iMarine catalogue of applications captures four main domain bundles that can be customized to meet specific needs:

* BiolCube is for practitioners working with species occurrence data and taxonomic profiles. It helps them to generate new knowledge from this data (e.g. modelling and analysing distribution data, comparing checklists, and producing maps).
* GeosCube is for practitioners dealing with geospatial information. It helps them to properly access, consume, and produce data. For example, it allows them to perform data processing tasks via Web Processing Service standard, as well as visualizing and publishing such data via state-of-the-art techniques.
* StatsCube is for practitioners working with a rich array of information, ranging from observational data to statistical data. It is a complete data-life-cycle-supporting framework, including data validation, data enrichment, and efficient analytical tools.
* ConnectCube is for practitioners wanting to produce information-rich objects, resulting from the aggregation and synthesis of data from multiple sources. It offers a comprehensive tool suite, which supports a collaborative, standards-oriented data-publication environment, including semantic technologies.

Rich set of capabilities

The iMarine data infrastructure is developed to support decision making in high-level challenges that require policy decisions typical of the ecosystem approach. For example, researchers who need to apply complex models to analyse multidisciplinary data. Managing the diversity of data formats and of temporal and spatial scales is one of the most significant challenges facing this community. The data infrastructure can also be applied to other high-level challenges like climate change or societal challenges.

By bringing together a rich set of capacities, iMarine provides an ideal platform supporting the development of partnerships among the knowledge institutions involved in the implementation of the Ecosystem Approach to Fisheries Management and conservation of living marine resources.

The iMarine infrastructure combines over 500 components into a coherent and centrally managed infrastructure of hardware, software, and data resources. gCube serves as the underlying software system to build and operate iMarine as a hybrid data infrastructure. Key technology enablers supporting open science include:

* A 'system of systems' based on the federation of existing systems, including grid- and cloud-based solutions, helping to create a self-contained infrastructure.
* Computing and storage capacities for the efficient and elastic provision of compute and storage services, scaling up and down to meet demand.
* Data management facilities as a service capable of tackling a rich array of data types, ranging from papers to tabular data and maps and a comprehensive set of research products. The service-oriented-architecture approach enables a building block approach to extending these facilities as new needs emerge.
* Virtual research environments offer flexible and secure web-based, community-centric platforms, so researchers can work together on common challenges. These collaborative platforms, as well as a complete set of accompanying tools, can be created quickly and easily.

iMarine integration and core services. Examples of iMarine integration of technologies and resources range from grid and cloud computational service provision to biodiversity, climate and environmental data access and through interoperability with resources such as GBIF (Global Biodiversity Information Facility), OBIS (Ocean Biogeographic Information System), and the FAO-FI reporting schema, among others. Other core iMarine services span data reg-

istration and access, data harmonisation validation and enrichment, data transformation, publishing and visualisation, analysis and providing collaborative platforms like virtual research environments.

The iMarine Gateway. As the key enabling technology, gCube software ensures seamless access to different technological resources, which together provide a hybrid data infrastructure. The iMarine Gateway provides a single point of access to specialised gCube applications with functionalities to manage, process and visualise scientific data. This common interface provides user communities with a comprehensive tool-box.

Virtual Research Environments. The iMarine data infrastructure has the capacity to quickly create virtual research environments as collaborative platforms connecting geographically dispersed user communities who want to work on a common goal. The **gCube Virtual Research Environment (VRE) technology** therefore enables a **user-centric** and **workflow approach**. The iMarine virtual research environment is a distributed and dynamically created environment, where the data, services, computational and storage resources are governed by dedicated policies. These services and resources are assigned to users via interfaces for a limited timeframe at little or no cost for the providers of the participatory data infrastructures.

This service-oriented approach improves not only legacy application quality of service without incurring large reimplementation costs but also application capabilities. Besides this approach, iMarine also leverages Platform-as-a-Service (PaaS), transforming complex platforms into a service, practitioners can benefit from data management capacities and information retrieval capacities on a ready-to-use basis.

Future horizons to tackle key challenges

On 14-15 May 2013, the European Commission hosted a high-level workshop on iMarine e-infrastructure for data-driven decision making and research. The workshop helped identify opportunities for iMarine to scale out its user base by establishing bilateral agreements and supporting policy-related services.

Mapping of ecosystems: including marine ecosystems: the Commission is facing challenges because the spatial databases are usually partial. There are different sea conventions and different projects dealing with different parts of the seas but we lack full coverage. The first task, led by the European Environmental Agency, is how to have an overview of what is going-on in the seas. The plan is to elaborate data to measure certain trends and understand environmental parameters (e.g. temperature). iMarine can help address this challenge by providing indicators.

Assessing the conditions of ecosystems: very much dependent on the quality of data coming from reporting obligations. From the directive on marine species and habitats, the EU assessment will be based on this year's reporting by the end of next year. On the water assessment, there are aspects related to marine which are covered but the main input will be on the reporting under the marine strategy framework directive this year. The European Environmental Agency is tasked with providing a first baseline assessment by next year. For the quality descriptors, there are 11 descriptors and this will be really a key component for the assessment of the quality of the seas but we will also look at other datasets especially from air pollution in order to see the impact or some pressures on the marine ecosystems.

Supporting ecosystem services - access issue for fisheries data, socio economic data and data from statistical offices: these issues are currently being discussed at European level. Requirements of the kind of data required have to be very precise. This activity is led by the Commission's Joint Research Centre (JRC). Integrating all this into a single, integrated assessment: developing evaluation methods for the ecosystem services and include the integration in accounting systems (by 2014).

Reducing data discrepancies - efforts to minimize discrepancies between scientific reporting and official fisheries data. A DG Mare objective is to reduce the risk of inaccurate data and misreporting. Simplifying workflows, creating a linear framework and standardization are important steps in mitigating these risks.

Supporting flows of controlled data - flows of controlled data can be used for science, whereas scientific data cannot be used, for example, to verify declared versus real catch statistics because inspection systems are used for this purpose. iMarine could support flows of controlled data to the scientific community under the provision of confidentiality requirement policies. It could also consider offering specific services that are currently missing in the policy space.

User-centric approach, support and web-based training facilities

iMarine is working to ensure that users are right at the very centre of its operations. It has segmented the market, matching its services and tools for different user communities. iMarine has also developed a stakeholder engagement plan that looks beyond the scientific community, such as NGOs, which can help extend its market reach. The iMarine web-based training environment can support uptake of services.

iMarine has identified several organisations and initiatives that could benefit from its computational capacity and approach to addressing data challenges, with the aim of:

* Scaling the data infrastructure resources to meet computational demand – one of the key needs identified during the workshop.
* Offering data services on behalf of the data provider by transforming the data and making it available in a standard format.

Benefits: Joining as a node to the data infrastructure means immediate access to monitoring and service support. iMarine monitors the availability of a node. If more resources are needed, it can scale up. It

therefore offers tangible benefits in terms of services on top of services for the validation of the metadata thanks to the data infrastructure and its design concept.

Sample of potential uses: The International Council for the Exploration of the Seas (ICES) has groups that are struggling with stock assessments. One of the ICES groups is using cloud solutions to speed up the time it takes to make a stock assessment, from a month to a week. There is an increasing transition to cloud uptake within ICES groups.

Practical classroom usage also offers new exploitation opportunities. GEOMAR has an interface that guides users through a step-by-step process for a complex analysis to estimate Maximum Sustainable Yields (MSY). This process takes several minutes for one user but GEOMAR lacks the capacity to handle demand. Currently, experiments are run on our servers sequentially, which takes about 40 minutes. GEOMAR lacks the capacity to handle demand. iMarine could provide meet this capacity need. In exchange, the GEOMAR interface could include a message saying "this service is powered by iMarine". Such an approach would offer access to thousands of users and give considerable visibility. Having an IP address to give to the users would be easier than registration.

In practical terms, iMarine has training courses on integrating models or can provide direct support. The classroom can become a collaborative platform through the creation of a virtual research environment with the allocation of the required resources.

Authors

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

Participants Workshop May 2013

European Commission: Antonios Barbas and Carlos Morais-Pires, e-Infrastructure Unit, DG CONNECT; David Connor, Marine Environment and Water Industry, DG Environment; Olivier Fontaine, Maritime policy for the Mediterranean and Black Sea, DG MARE; Amelie Knapp, Structural action: Ireland, Spain, France, Portugal and the UK, DG Maritime Affairs and Fisheries; Anne Teller, Biodiversity Policy Officer, DG Environment; Waddah Saab, DG RTD.

EU agencies: Darja Lihteneger, European Environmental Agency (EEA).

EU & global affiliations: Maud Evrard, Science Europe; Rainer Froese, GEOMAR; Giuseppe Manzella, ENEA; Yolanda Sagarminaga, AZTI Tecnalia.

EC programmes & global initiatives represented: European Marine Observation and Data Network – Biology and Physics (EMODnet); FishBase; JPI Oceans; EurOBIS.

National government agencies & public institutions: Pablo Aduanza, Department of Fisheries, Spanish Institute of Oceanography

iMarine representatives: Nicolas Bailly, FIN ; Donatella Castelli, Gianpaolo Coro and Pasquale Pagano, ISTI-CNR; Willem Demoor, VLIZ; Anton Ellenbroek and Marc Taconet, FAO Fisheries and Aquaculture Department, Fisheries Information and statistics Branch (FIPS); Sara Garavelli and Silvana Muscella, Trust-IT Services; Panagiotis Liakos, Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens.

Members of the iMarine Board represented include (in alphabetical order): Ward Appeltans, Ocean Biogeographic Information System (OBIS) - Intergovernmental Oceanographic Commission of UNESCO; Patrico Bernal, IUCN High Seas Initiatives Coordinator; Ingrid Bergeret, Head of Fisheries and Aquaculture Information System, French Ministry of Ecology, Sustainable Development and Energy (DPMA/MSIPA); Francky Callewaert, Integrated Fisheries Data Management, DG Mare; Mark Dickey-Collas on behalf of Neil Holdsworth, International Council for the Exploration of the Sea (ICES); Rachel Lewsley, North East Atlantic Fisheries Commission.