



EGI-Engage

D2.7 Market Report on the Fishery and Marine Sciences Data Analysis Sector

D2.7

| | |
|------------------------|---|
| Date | 13 June 2016 |
| Activity | NA2 |
| Lead Partner | ENG |
| Document Status | FINAL |
| Document Link | https://documents.egi.eu/document/2700 |

This document introduces the Fishery and Marine Sciences Data Analysis Sector to EGI through a top-down approach, which presents a market analysis, a domain and stakeholder analysis, and data flow analysis. In addition, a bottom-up approach was adopted and presents a survey analysis (questionnaire, interview, example initiative serving the community).

The work is supported by EGI-Engage TNA2.3, and works towards its objectives by developing the EGI business engagement programme and service exploitation programme by offering services for exploitation of fishery and marine data to innovators. The expected outcome is new collaborations with SME's and other stakeholder types. This will be achieved by providing a clear introduction to the sector, its domains, stakeholders, and data value chains. Furthermore, this document allows EGI to learn the data needs and challenges of the different stakeholder types, giving insights about the community requirements, preparing EGI for any future engagement/collaboration (if desired) which could lead to the creation of future business collaborations.



This material by Parties of the EGI-Engage Consortium is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

The EGI-Engage project is co-funded by the European Union (EU) Horizon 2020 program under Grant number 654142 <http://go.egi.eu/eng>

COPYRIGHT NOTICE



This work by Parties of the EGI-Engage Consortium is licensed under a Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>). The EGI-Engage project is co-funded by the European Union Horizon 2020 programme under grant number 654142.

DELIVERY SLIP

| | <i>Name</i> | <i>Partner/Activity</i> | <i>Date</i> |
|----------------------|----------------------|-------------------------|-------------|
| From: | Nadia Nardi | ENG | 31.01.2016 |
| Moderated by: | Małgorzata Krakowian | EGI.eu/NA1 | |
| Reviewed by | Tiziana Ferrari | EGI.eu/PMB | 02.03.2016 |
| | Peter Solagna | EGI.eu/AMB | 22.02.2016 |
| | Bartosz Kryza | AGH UST | 18.02.2016 |
| Approved by: | AMB and PMB | EGI | 03.03.2016 |

DOCUMENT LOG

| <i>Issue</i> | <i>Date</i> | <i>Comment</i> | <i>Author/Partner</i> |
|--------------|-------------|---|--|
| v.1 | 10.12.2015 | First draft, integration of contributions | Nadia Nardi, ENG Yann Laurent, FAO Anton Ellenbroek, FAO |
| v.2 | 29.01.2015 | First full version of content for internal review | Nadia Nardi, ENG Yann Laurent, FAO Anton Ellenbroek, FAO Pasquale Pagano, CNR |
| v.3 | 18.02.2016 | Integration from external review | Nadia Nardi, ENG |
| v.4 | 23.02.2016 | Integration of review from AMB | Nadia Nardi, ENG |
| v.5 | 02.03.2016 | Final version | Nadia Nardi, ENG |
| v.6 | 07.06.2016 | New final version to implement EU reviewer comment (Minor typographical errors) | Nadia Nardi, ENG |

TERMINOLOGY

A complete project glossary is provided at the following page: <http://www.egi.eu/about/glossary/>

ACROYMNS

| Acronym | Description |
|----------------|--|
| AIS | Automatic Identification System |
| CFP | Common Fisheries Policy |
| CPUE | Catch per Unit Effort |
| DaaS | Data as a Service |
| EC | European Commission |
| EEZ | Exclusive Economic Zone |
| EMFF | European Maritime and Fisheries Fund |
| EwE | Ecopath with Ecosystem |
| FAO | Food and Agriculture Organization |
| FTE | Fully Time Equivalent |
| GDP | Gross Domestic Product |
| GIS | Geographic Information System |
| ICES | International Council for the Exploration of the Sea |
| IOTC | Indian Ocean Tuna Commission |
| ISO | International Organization of Standardization |
| IUU | Illegal, Unreported and Unregulated Fishing |
| MCS | Monitoring, Control and Surveillance |
| MPA | Marine Protected Area |
| MSC | Marine Stewardship Council |
| NATO | North Atlantic Treaty Organization |
| REC | Regional Economic Commission |
| RFMO | Regional Fisheries Management Organization |
| SME | Small Medium Enterprise |
| VMS | Vessel Monitoring System |

Contents

| | | |
|-------|--|----|
| 1 | Executive Summary | 6 |
| 2 | Introduction..... | 8 |
| 3 | Background..... | 10 |
| 3.1 | Objectives | 10 |
| 3.2 | Scope | 10 |
| 3.3 | Intended Audience | 10 |
| 3.4 | Approach..... | 10 |
| 4 | Landscape/Seascape..... | 12 |
| 4.1 | Market Overview | 12 |
| 4.2 | Fishery and Marine Sciences Domains | 13 |
| 4.2.1 | Domains Included in the Study | 13 |
| 4.2.2 | Domains Excluded from Further Study | 19 |
| 5 | Stakeholder Analysis..... | 21 |
| 5.1 | Identification of Stakeholders | 21 |
| 5.2 | Data Dimension..... | 22 |
| 5.2.1 | Data Owners | 22 |
| 5.2.2 | Data Processers | 23 |
| 5.2.3 | Data Consumers..... | 23 |
| 5.3 | Stakeholder Characterization | 23 |
| 5.4 | Stakeholder Mapping- Data Value Chain..... | 25 |
| 5.4.1 | Fisheries Data Value Chain..... | 25 |
| 5.4.2 | Aquaculture Data Value Chain | 27 |
| 5.5 | Data Flows..... | 28 |
| 5.5.1 | Fisheries Exploitation and Monitoring Domain | 29 |
| 5.5.2 | Fisheries Catches: Traceability/Certification/Quality Control Domain | 31 |
| 5.5.3 | Marine Fisheries Research Domain..... | 33 |
| 5.6 | Domain and Stakeholder Findings Representing Opportunities for EGI | 34 |
| 6 | Survey Analysis..... | 36 |
| 6.1 | Approach and Objectives | 36 |
| 6.2 | Analysis | 36 |

| | | |
|-------|--|----|
| 6.2.1 | Questionnaire | 36 |
| 6.3 | Interviews | 50 |
| 6.4 | The BlueBRIDGE Case | 51 |
| 7 | Findings, Conclusions and Next Steps | 54 |
| 7.1 | Summary of Key Findings | 54 |
| 7.2 | Conclusions and Next Steps | 61 |
| 8 | References..... | 62 |
| | Appendix A | 63 |

Figures table

| | |
|--|----|
| Figure 1 - Fisheries Data Value Chain | 26 |
| Figure 2 - Aquaculture Data Value Chain | 27 |
| Figure 3- Fisheries Exploitation and Monitoring Domain- Data Flow | 29 |
| Figure 4- Fisheries Catches: Traceability/Certification/Quality Control- Data Flow | 31 |
| Figure 5- Marine Fisheries Research Domain Data Flow | 33 |
| Figure 6 - Question 8 - What is the current scope of the data you manage? | 38 |
| Figure 7 - Question 9 - What type of data does your institution/company manage?..... | 39 |
| Figure 8 - Question 10 - Select the top three priorities for your institution/company? | 39 |
| Figure 9 - Question 13 - In 2-5 years from now, what do you believe will be your main challenge regarding the management of data? | 42 |
| Figure 10 - Question 14 - Does your institution/company use Cloud-based services?..... | 43 |
| Figure 11 - Question 15 - If your institution/company uses Cloud-based services, which services?..... | 43 |
| Figure 12 - Question 16 - If your institution/company uses Cloud-based services, what type of Cloud?..... | 44 |
| Figure 13 - Question 17 - If your institution/company is already using some Cloud-based Services, do you plan to further invest in other services? If yes, which ones?..... | 44 |
| Figure 14 - Question 18 - If your institution/company is not using Cloud-based services although is interested, what services would be of interest within the next two years?..... | 45 |
| Figure 15 - Question 19 - Are you aware of any obstacles/challenges your institution/company recognizes in regards to using Cloud-based services? If yes, in which areas? | 46 |
| Figure 16 - Question 21 - On a scale from 1 to 4, how important is the Cloud computing benefit: Facilitate analysis of cross-domain information through harmonization and standardization..... | 47 |
| Figure 17 - Question 22 - On a scale from 1 to 4, how important is the Cloud computing benefit: Cost saving..... | 48 |
| Figure 18 - Question 24 - On a scale from 1 to 4, how important is the Cloud computing benefit: Performance and flexibility | 49 |

1 Executive Summary

The purpose of this document is to report on the business opportunities of EGI in providing Data as a Service (DaaS) – combining access to relevant core datasets, computing and applications, in the Marine and Fishery Data Analysis Sector. DaaS builds on the concept that the data can be provided on-demand together with the necessary analysis facilities to the user regardless of geographic or organizational separation between provider and consumer.

EGI has just begun to collaborate with the Fishery and Marine Sciences community by requesting this study and also through two other Project Tasks: firstly, the integrating of iMarine Virtual Research Environments¹ (the gCube/D4Science platform)² and their data with the EGI Federated Cloud, and secondly, the go-to-market stage with the operations of the Fishery and Marine Sciences VREs. This activity allows for the development of the business proposition of EGI by augmenting the community-specific PaaS enabled by the EGI cloud platform.

This document contributes to prepare EGI for any future engagement that could lead to the creation of future business projects. This document, supported by the Task NA2.3, facilitates the connection of EGI with the stakeholders by providing a clear introduction to its domains³ and peculiarities. Additionally, this document provides an informed analysis of the data requirements and obstacles expressed by the stakeholders themselves.

The approach taken for this document includes both a top-down and bottom-up analysis. The top-down approach consists of a market overview that gives an idea of the structure and size of the sector through facts and figures. The sector is described in terms of domains and serves as a presentation to what exactly the Fishery and Marine Sciences Data Analysis Sector is about, and the respective activities/purposes. Stakeholders are analysed, identifying and characterizing the stakeholders through different dimensions, this to familiarize the reader with what type of data is managed and in which way. Specifically, stakeholders are described in terms of:

- the type of entity (public/private);
- whether they are monitoring and managing organizations or exploitation companies;
- the type of data used (detailed, aggregated data/statistics, reference/standard/classification data);
- the role held in the data value chain: owners, processors or consumers;
- the activities they take forward, the purpose and scope of the data they work with.

Additionally, stakeholders are mapped to two data value chains: one for Fisheries and the other for Aquaculture, this to give a visual representation as an overview. Finally, stakeholders are mapped to the two main sub-sectors: Fisheries and Aquaculture, where the flow of data and information (from collection to exploitation) in three chosen domains is documented.

¹ <http://www.i-marine.eu/Pages/Home.aspx>

² <https://www.d4science.org/>

³ The word “domain” is used to break-down the Fishery and Marine Sciences Data Analysis Sector into sub-sectors, rather fields of study/action.

The bottom-up analysis consists of the analysis of the results of a survey organized to give the reader other information directly collected from the various stakeholders in the domains of interest. This is done through the analysis of a distributed questionnaire. The purpose of the questionnaire was to understand the computing needs and obstacles encountered by the stakeholders. Questions, both closed and opened were asked about data management aspects and cloud-based services, also considering a future outlook involving EGI. Additionally, to collect other qualitative information, interviews were performed with selected data managers. Finally, in order to include a different perspective, a successful example of an initiative that has been serving this community for many years is described.

The overall findings of this study are positive. About half of the respondents use cloud-based services, and others are interested in or plan to move to cloud-based provisioning in the next 2-5 years. This surely gives EGI an indication that the community is ready for services and solutions that EGI has to offer, and perhaps would be interested in co-creating. Only one of the respondents is familiar with EGI. This is indeed new territory for EGI, where every finding is useful to learn about the community. One of the most important findings is the interest in a new service combining different types of cloud services, besides storage and computing infrastructure, also the reuse of data from third parties, software, in addition to technical support.

Furthermore, another notable Finding, confirmed through the questionnaire, is the need on behalf of the respondents to access external data, not only from other domains in the Fishery and Marine Sector, but also in other Sectors and related domains, such as geospatial and biodiversity. Access to relevant data in these “other” domains was rated the second highest priority for the respondents. Data-as-a-Service (DaaS) could certainly appeal to the Fishery and Marine Sciences Sector.

Moreover, the Findings meet the deliverable intention of producing valuable input to be further analysed in collaboration with the EGI User Community Support and Operations Teams. This analysis will ensure the identification of technical requirements (in the most appropriate format) to take forward and make use of, also through a business analysis for the integration or creation of new services.

The top-down approach coupled with the bottom-up approach for a market report of this nature is an effective way to introduce the Fishery and Marine Sciences Data Analysis Sector, and give valuable insights into a community relatively unknown to EGI.

2 Introduction

EGI-Engage⁴ aims to accelerate the implementation of the Open Science Commons⁵ by expanding the capabilities of a European backbone of federated services for compute, storage, data, communication, knowledge and expertise, complementing community-specific capabilities. One of its objectives is to evolve EGI services and solutions portfolio, and related business models according to customer needs. This document works towards this objective through WP2 (NA2) Strategy, Policy and Communications⁶, looking to steer the consolidation and growth of the EGI community by developing a strategy towards the Open Science Commons vision and to ensure the engagement of the all stakeholders; in this case, the Fishery and Marine Sciences Data Analysis Sector.

This document is delivered under the Task NA2.3 SME/Industry Engagement and Big Data Value Chain. It seeks to:

1. Facilitate the connection of EGI with SMEs and other stakeholder types at a European level by providing a clear introduction to the sector, its domains, and stakeholders.
2. Allow EGI to learn the data needs and challenges of the different stakeholder types in the Fishery and Marine Sciences Data Analysis Sector. This will give EGI insights on the community requirements and in turn help EGI in the creation of enhanced services unifying computing and data approaches.

In contributing to the project aims of EGI-Engage, this document constitutes as an input to the mission of EGI-Engage, should it be interested to further investigate and work with the Fishery and Marine Sciences Data Analysis Sector, also for the creation of a model for stakeholder engagement. Additionally, this document provides pertinent insights for EGI to be able to attract the stakeholders to collaborate and then to explore and detect opportunities and threats around the Open Data, and co-develop business models for their exploitation.

This document is important to EGI-Engage as it introduces EGI to the Fishery and Marine Sciences Data Analysis Sector. EGI should be interested in this community because there are opportunities to be taken. To note, EU investment in this sector is not indifferent, the European Maritime and Fisheries Fund (EMFF)⁷ has a budget of around €5.749 billion for the period 2014-2020. The budget is allocated to support improving fisheries data collection, and to allow decisions to be based on robust evidence, amongst other areas of concern. The data this sector manages are huge, and range from on-board monitoring of bycatch⁸, to economic and employment indicators.

⁴ <https://www.egi.eu/about/egi-engage/>

⁵ http://www.egi.eu/news-and-media/publications/OpenScienceCommons_v3.pdf

⁶ <https://wiki.egi.eu/wiki/EGI-Engage:WP2>

⁷ http://ec.europa.eu/fisheries/news_and_events/events/national_strategic_plans/emff_en.pdf

⁸ A fish or other marine species that is caught unintentionally while catching certain target species and target sizes of fish, crabs etc.

This document begins with a chapter named Background, where information covering document objectives, what is in and out of scope, which is the intended audience, and the overall approach taken forward to produce the market report.

To follow is a chapter on the Landscape/Seascape of this sector. This includes a market overview that gives an idea of the Sector's structure and size of the sector. The sector is then described in terms of domains and serves as a presentation to what exactly the Fishery and Marine Sciences Data Analysis Sector is about and the respective activities/purposes. A special emphasis is given to the domains that may be of higher interest to EGI.

To follow is a Chapter on Stakeholder Analysis. It identifies and characterizes the stakeholders with different dimensions. This is to familiarize the reader with what type of data is managed and in which way. Additionally, the stakeholders are mapped in a data value chain; one for Fisheries and other for Aquaculture. The chapter ends with a section presenting data flows on three chosen domains.

Furthermore, the next chapter, Survey Analysis, aims to give the reader other information directly collected from the various stakeholders in the domains of interest. This is done through the analysis of the distributed questionnaire and performed interviews with selected data managers. The chapter ends by describing a successful example of an initiative that has been serving this community for many years.

Finally, the document ends with a chapter on the findings and recommendations for EGI to easily discover.

3 Background

3.1 Objectives

This study represents an introduction of the Fishery and Marine Sciences Data Analysis Sector to EGI, as this community is relatively new to EGI. However, there is an expressed interest to understand the community's data needs for possible future relationships. Through a top-down market analysis, which documents market size and structure and explores possible high interest domains, EGI will gain insights into SME, industry, and academia data requirements. In addition, a bottom-up analysis was conducted through questionnaires distributed to stakeholders, and interviews performed with data managers. This may facilitate EGI to connect with the community, having also a qualitative analysis at hand.

3.2 Scope

For clarity, when referring to the Fishery and Marine Sciences Data Analysis Sector, the Fisheries⁹, Aquaculture¹⁰ and Maritime sectors are included.

A description of the main domains in the sector is given below in Section 4.2, followed by the identification of which domains the study will concentrate on. The domains chosen are those that could be of high interest to EGI.

3.3 Intended Audience

The intended audience is the EGI community and external experts (e.g. RDA¹¹, BlueBRIDGE¹² initiative) who are interested in an introduction to the Fishery and Marine Sciences Data Analysis Sector and/or in the data needs of the stakeholders expressed by the stakeholders themselves.

3.4 Approach

The approach chosen is to introduce the reader to the Fishery and Marine Sciences Data Analysis Sector through a top-down and bottom-up analysis. The top-down analysis includes a desk study (market analysis) describing the size and structure of EGI's potential market through facts and figures, and information about EC funding. Key materials include the FAO's 2014 The State of

⁹ Fishery <http://www.fao.org/faoterm/viewentry/en/?entryId=98327>

¹⁰ Aquaculture <http://www.fao.org/faoterm/viewentry/en/?entryId=1222>

¹¹ <http://europe.rd-alliance.org/rda-europe>

¹² <http://www.bluebridge-vres.eu/>

World Fisheries and Aquaculture¹³ and the EC's 2014 Facts and figures on the Common Fisheries Policy¹⁴. Moreover, the Fishery and Marine Sciences Data Analysis Sector is explained through the description of the domains found under its umbrella; particular attention is given to high interest domains. A detailed stakeholder analysis and the documentation of data flows reveal insights specific to these communities. Furthermore, the bottom-up analysis will complete the picture by giving insights on the data needs directly from the stakeholders (data managers) of the interested domains through the analysis of the questionnaires they responded to and interviews held with them. Also, the project BlueBRIDGE is presented as an initiative that successfully serves the Fishery and Marine Sciences Data Analysis domains.

¹³ <http://www.fao.org/3/a-i3720e/index.html>

¹⁴ http://ec.europa.eu/fisheries/documentation/publications/pcp_en.pdf

4 Landscape/Seascape

This Chapter begins with a brief overview of the size of the Fishery and Marine Sciences Data Analysis Sector through different facts and figures. To follow is a comprehensive introduction to the domains in the Fishery and Marine Sciences Data Analysis Sector. Particular focus is given the domains that may be of high interest to EGI.

4.1 Market Overview

Ocean fisheries add \$270 billion to global GDP¹⁵, providing protein for nearly 3 billion people. The FAO estimates that, overall, fisheries and aquaculture assure the livelihoods of 10–12 percent (260 million) of the world's population. Capture fisheries account for some 90 Million tonnes annually.

The EU has an important presence with a combined fleet in 2014 of around 87,000 vessels, together generating an income of €7.2 billion, and employment reaching 110,000 FTE's¹⁶. Aquaculture accounts for about 20% of fish production and directly employs some 80,000 people¹⁷.

The average EU citizen consumes 23.1 kg of fish products per year, and there is an ever-increasing need to know from where and how the fish arrived at the consumer.

The data needs to manage this sector are huge, and range from on-board monitoring of bycatch, to economic and employment indicators.

The EU Structural policy in the fisheries sector contributes to the objectives of the Common Fisheries Policy (CFP)¹⁸. The financial instrument of this policy is the European Maritime and Fisheries Fund (EMFF)¹⁹, which has a budget of around €5.749 billion for the period 2014-2020. The budget is allocated to support improving fisheries data collection, and to allow decisions to be based on robust evidence, amongst other areas of concern.

¹⁵ The State of World Fisheries and Aquaculture, 2014, <http://www.fao.org/3/a-i3720e/index.html>

¹⁶ EU-DGMARE http://ec.europa.eu/fisheries/documentation/eu_fisheries_key_facts/index_en.htm

¹⁷ DGMARE Aquaculture employment http://ec.europa.eu/fisheries/cfp/aquaculture/index_en.htm

¹⁸ http://ec.europa.eu/fisheries/cfp/index_en.htm

¹⁹ http://ec.europa.eu/fisheries/news_and_events/events/national_strategic_plans/emff_en.pdf

4.2 Fishery and Marine Sciences Domains

The Fishery and Marine Sciences Data Analysis Sector is comprised of many domains. Below is a description of a good part of these. The list is non-exhaustive, but the domains chosen were to give a good introduction to what exactly the Fishery and Marine Sciences Data Analysis Sector is about. Of the domains described, this report will focus on those that may be of higher interest to EGI, as they present opportunities. A special emphasis will be given to these domains.

4.2.1 Domains Included in the Study

The five domains below are where large volumes of data in Fisheries are found:

1. Marine Fisheries Exploitation and Monitoring, especially for industrial fisheries
2. Marine Fisheries Research to provide stock assessment (few types of data but with large volume)
3. Fisheries Catches Traceability/Certification/Quality Control (large number of different types of data from different stakeholders)
4. Marine Environmental Research. The trend is to compare data across domains, for example the impact of fisheries versus impact of tourism on the ecosystem. There is a need for harmonized data by public companies but also private sector consultancy companies.
5. Marine Aquaculture Research

Below is the description of nine domains (including the five above), which were chosen to be furthered explored, as they represent communities where EGI may find opportunities to serve.

4.2.1.1 *Marine Fisheries Exploitation and Monitoring*

Marine Fisheries exploitation encompasses all human activities related to marine species catch:

1. Fishing vessel building;
2. Registering for sailing and fishing licenses;
3. Sailing to fishing zone(s) once authorized;
4. Carrying out fishing operation²⁰ (can be over several weeks);
5. Sailing back to port to land catches;
6. Carrying out fishing vessel maintenance activities.

Marine Fisheries Monitoring aims to record all Marine Fisheries Exploitation activities to compute statistics on:

1. How many vessels fish a given species (Fishing effort)
2. How many fish are caught during these activities²¹ (Nominal Catches and Landed Weight.)

²⁰ Targeted marine species are broad, from small grey shrimps to large sharks, from bottom species (demersal fish like grouper or cod) to highly migratory species (large pelagic fish such as red tuna).

The challenge with fisheries is the uncertainty of fish stock status. Unlike agriculture or aquaculture, which are animal production activities, marine fisheries are wild animals harvesting activities. The knowledge of the quantity of fish to be taken out by fishermen is crucial for stock exploitation sustainability.

Large quantities of data are collected to monitor the stock exploitation level, especially for industrial fisheries. Combined with stock assessment coming from Marine Fisheries Research fisheries management policies can be made to ensure a sustainable level of fish stock exploitation. This domain is of interest as main data providers in fisheries have different challenges in terms of data storage, processing and dissemination.

Traditionally, the difference is made between industrial fisheries (commercial activities with large vessels) and small-scale or artisanal fisheries (for local sales or subsistence with smaller boats, usually undecked). The difference has a strong impact on data collection, industrial vessels landing in ports (data collection can be organized easily – industrial vessels are more and more equipped with electronic reporting tools such as e-logbook) although artisanal vessels can land literally anywhere on the coast even if main landing sites are known: catches monitoring is more costly as all main landing sites should be in theory monitored, although sample based surveys can be put in place for cost efficiency if local fisheries statistician resources are available.

4.2.1.2 *Marine Aquaculture Production and Monitoring*

Aquaculture in terms of production methods is closer to agriculture livestock than marine fisheries. Aquaculture production is organized in farms. To simplify, aquaculture produces fingerling²² after fish reproduction, grow fingerling to adult fish, and then grow the adult to the size adapted to the targeted market. In some cases, aquaculture is limited to fingerling or small adults growing like for red tuna.

Marine aquaculture is mainly done offshore in large circular cages. Main cultured species are salmon (*Salmo salar*²³) and molluscs²⁴. In comparison, freshwater aquaculture is much more developed than marine aquaculture, especially in China (3/4 of world aquaculture production is Chinese carps).

Marine aquaculture development has huge potential once high value species like lobsters and red tuna will be completely managed (breeding is still a problem for red tuna).

Main challenges for marine aquaculture are:

²¹ Nominal catches and landed weight are standard denominations for fisheries exploitation indicators as defined by the Coordinating Working Party on Fisheries Statistics (CWP) <http://www.fao.org/fishery/cwp/en>

²² Fingerling definition <http://www.fao.org/faoterm/en/?defaultCollId=14>

²³ https://en.wikipedia.org/wiki/Atlantic_salmon

²⁴ <https://en.wikipedia.org/wiki/Mollusca>

1. Competition with tourism for farm implementation (cages need still or protected waters, well suitable to develop tourism activities especially in warm waters like in Mediterranean sea)
2. Competition with wild animals, with a risk of contamination of wild species with selected domestic species (especially if the domestic one has been genetically modified) and with a risk of the transmission of disease from cultured animals to wild ones.

Production monitoring is done through administrative records or census. Production quantity has a direct impact on environmental pollution. It is usually monitored on site by public officer. A number of farms and total production for a given zone provide elements for aquaculture production policy at local and national levels. Monitoring through satellite images is currently being studied.

This domain is of interest as main data providers in aquaculture have different challenges in terms of data storage, processing, dissemination and innovation for data monitoring.

4.2.1.3 Fisheries/Aquaculture Catches - Traceability/Certification/Quality Control

The aim of certification is to ensure to the final customer that the sea product he/she buys comes from a sustained fishery. Certification is guaranteed by certificates delivered by authoritative organizations such as Marine Stewardship Council²⁵ (MSC). It relies on standard measurements and information to certify fish and fishery.

Certification consumes data coming from different levels:

1. Certification requires traceability of sea product: where and how it has been caught, where it has been landed, where it has been bought, where and how it has been processed. A unique identifier has to follow the fish lot all along the supply chain from vessel to customer plate.
2. Certification also requires proof that the fished stock is sustainably managed: information and data are sent by national fisheries institutions to certification institutions on stock management.
3. Certification requires that all sanitary measures are in place for the most efficient fish preservation all along the fish supply chain (cold chain preserved on board, at dock, in sale house, in fishmonger, no use of spoiled water to clean fish, etc.): controls/declarations are made at each level and this information must follow the fish lot for traceability.

²⁵ <https://www.msc.org>

In terms of data collection/processing, the key elements here are the definition of a unique identifier for landings and the capacity to exchange data between institutions and private sector to keep this traceability. This domain is of interest as it is a domain requiring an important volume of data discovering and exchange.

In regards to the capacity to exchange data between institutions, the level of maturity of the process is different depending on the case:

- In developed countries it works well, such as in France, UK, Germany etc...
- In developing countries, data exchange between stakeholders is an issue.
- Exchanges of data between institutions of different countries, for example Spain to UK or UK to France is more complex, as definitions and classifications can vary, but it happens.
- Exchanges from countries to regional/international organizations, it also works as standards are defined by the international organizations (although not all countries report to regional/International organizations - it works better when dealing with species of high value, such as tuna.)

DG MARE is addressing this issue of multi-reporting with the Integrated Fisheries Data Management Program both by defining standard formats to exchange (to be endorsed by UN/CEFACT) and the pipe to convey the data (with pieces of software to send and receive messages.)

The same applies to aquaculture. Certification validates that aquaculture implementation is respectful of natural ecosystems, for example, mangroves. Mangroves are not destroyed to implement fish farms, spoiled water coming from shrimp ponds are cleaned in specific areas and not directly rejected in the sea, etc.

4.2.1.4 *Marine Fisheries Research*

Marine Fisheries Research conducts a broad variety of studies on fisheries, which can be roughly classified in five sub-domains:

1. Species (including those of economic interest) studies: description, biology and interactions between species (e.g. sperm whales eating anchovies in competition with fishermen / fish population dynamic in ecosystem)
2. Stock population estimation (e.g. population modelling for a given species in a given area, fish population)
3. Stock assessment: integration of biological data on a given species and level of catches/landing for the species in a given area: provide recommendation on the stock status (moderately exploited, fully exploited, overexploited, depleted, recovering, etc.) and adapted fisheries management measures (quotas, ban of zone or gear type, creation of Marine Protected Areas, etc.)

4. Improvement of fisheries methods (e.g. decrease by-catch mortality, new methods such as Catch Aggregating Devices, etc.)
5. Impact of external activities on fisheries (impact of oil activities/sea wind turbines/piracy on fishing activities, etc.)

Research data type reflects the large variety of fisheries related studies. Large amount of data can be collected in different formats (data can be stored in Excel tables, in databases such as PostgreSQL (very popular in research) or as R tables (R being an open source statistical processing tool)) and from various sources. Ad-hoc formats defined at the study level are implemented. Sharing this information would require harmonization and standardization.

A key element in research is the ability to store all data as it could be reused later on to re-validate the study itself and to exploit new methodologies (on stock assessment).

This domain is of interest as Marine Fisheries Research is a domain with large quantities of produced data in scattered projects that could benefit from cloud storage solutions.

Note: Usually researchers do not share their raw data. This presents a problem, because when they do, work has to be done to document raw data for exchange. Exchange is basically sharing the existing data in the existing format to another researcher. As a methodology, this is not very efficient as there is no common format to exchange data between researchers.

4.2.1.5 *Impact of Fisheries on Ecosystem Study – Marine Protected Areas (MPA) Efficiency*

These research activities are a sub-set of the Marine Fisheries Research domain (described above) but are important enough to be considered on its own.

Research is conducted to assess the impact of fisheries management measures on fish population recovery or protection. The number of Marine Protected Areas has increased dramatically in the past years with different targets in terms of percentage of MPA to cover ocean (10%-30% of whole oceans depending in the Conventions). Different types of MPA have been implemented from a total ban of fishing-to-fishing co-management by the MPA communities.

The challenge is the assessment of effect of this MPA and more broadly speaking of any fishing activity on ecosystem.

Large quantities of ad-hoc data are collected for such purposes.

4.2.1.6 *Maritime Control Surveillance/Safety at Sea*

Monitoring, control and surveillance (MCS) encompasses all activities related to monitoring, control and surveillance fishing activities. It also covers safety at sea control, looking to understand whether fishermen are trained and licensed to go to sea, if the vessel is sufficiently equipped with life vests and other similar safety equipment.

Different types of operations can be carried out to control fishing activities compliance with national, regional and international regulations. Within the Exclusive Economic Zones (EEZ²⁶), national authorities such as Coast Guards, air, navy or defence forces conduct maritime or air operations to monitor fishing activities. Regulation offences and infringements are detected such as fishing with expired license, fishing during a closed season, with unauthorized gear, in a closed area, etc. Offenders can be fined and the ship can be retained. Additional controls from shore can be done by radar control or AIS/Vessel Monitoring System (VMS)²⁷ control.

MCS requires good coordination between maritime authorities and fisheries authorities issuing licenses and fishing regulation with Navy or Coast Guards enforcing regulation. A list of authorized vessels must be exchanged before MCS operations start. VMS data can be obtained from different sources (private companies for instance). It implies data exchange according to standards.

The challenge in MCS is data processing coming from different sources.

4.2.1.7 Marine Fisheries Policy Making/Management

Marine fisheries management and policy making is the domain “on top” of the other marine fisheries related domains. It exploits information from marine fisheries exploitation, marine fisheries research and MSC to draft fisheries policies.

From technician and specialist recommendations, policy makers will define best suitable fisheries management policies for their countries, or for the regional organizations like the European Commission or Regional Fisheries Management Organization (RFMO) like the Indian Ocean Tuna Commission (IOTC), ICES²⁸, ICCAT²⁹) or the international one (Food and Agriculture Organization (FAO)). Fisheries Acts are amended, additional regulations can be voted and recommendations are endorsed/voted by Member States.

One outcome of making policy from different sources of information is to ensure comparability of these sources by imposing standards to data providers. The EC is in the process of standardizing fisheries reporting formats from all stakeholders (Member States, RFMO, etc.) through UN/CEFACT standards.

It will be interesting to assess the needs of RECs like the EC to use e-Infrastructures to centralize and store data for policy making purposes, and to distribute UN/CEFACT based software and standards.

²⁶ <https://ec.europa.eu/growth/tools-databases/minventory/content/eez>

²⁷ http://ec.europa.eu/fisheries/cfp/control/technologies/vms/index_en.htm

²⁸ <http://www.ices.dk/Pages/default.aspx>

²⁹ <https://www.iccat.int/en/>

4.2.1.8 *Marine Aquaculture Research*

Recent critics of the aquaculture production method include:

- High pollution level generated by farms
- Competition with wild animals and decreasing its genetic pool
- Spreading diseases to wild animals
- Use of wild fish meal in aquaculture food

Marine aquaculture research has several aims to address the above critics:

- Improvement of production methods (decrease of antibiotics use, decrease of escaped cultured fish into the wild polluting wild genetic pool)
- Improvement of cultured fish nutrition (substitution of fish meal with vegetal products, soya, cereals, etc.)
- Limitation of pollution (improve food distribution, food pellet composition)
- Improvement of cultured species (selection, genetically modified species to improve resistance to diseases)

Compared to Marine Fisheries, less data is collected, as there is no need for stock assessment and the complex related data collection of stock information, but there is a similar need for a large volume of data storage and processing.

4.2.1.9 *Marine Environmental Research*

This research domain aims to have broader views on marine environment than marine fisheries research. Hence, it covers a large variety of topics, such as:

- Taxonomic studies
- Biodiversity studies
- Impact of global warming on oceans' ecosystem
- Impact of tourism activities on coastal sea life (tourisms in sea turtles breeding area)
- Impact of human pollution on aquatic life
- Impact of offshore wind turbines on aquatic life and sea-birds

Some collected data series are really huge (temperature/salinity measures taken 5 times a day for 100 marine sites for 5 depths over 3 years: more than five million records to store and to process) and would benefit from big data capacities.

4.2.2 *Domains Excluded from Further Study*

The following six domains (listed here below) are part of the Fishery and Marine Sciences Data Analysis Sector. Although, they are not further studied in other sections, they are described in Appendix A to complete an overall view of the Sector. They are excluded from the study as when

taking into account the objectives of EGI-Engage to explore this sector for high interest areas to collaborate in the future; they were less attractive. Motivations include lower volumes of data compared to other domains, fairly new domain, and engagement possibilities for EGI.

1. Illegal, Unreported and Unregulated Fishing (IUU)
2. Coastal - Maritime Tourism
3. Oil / Gas Exploration and Exploitation and Infrastructure (Platforms, Pipes, Storage)
4. Maritime Transportation and Infrastructure (Main Sea Routes, Ports and Lighthouses)
5. Offshore wind turbines and related sustainable energy infrastructure (cables)
6. Piracy

5 Stakeholder Analysis

This chapter identifies the stakeholders of the nine domains of interest. The stakeholders are characterized through different dimensions: entity type, scope of data they work with, type of data, activities performed, and purpose/interest in the data. To give an overall view of Fisheries and Aquaculture, data value chains are also presented. Moreover, three domains were chosen to be explored in further detail, this to focus on the flow of data through the different entities. The chapter then includes initial insights and preliminary findings.

5.1 Identification of Stakeholders

The nine domains of interest, as detailed in Chapter 4 are:

1. Marine Fisheries Exploitation and Monitoring
2. Marine Aquaculture Production and Monitoring
3. Fisheries/Aquaculture Catches - Traceability/Certification/Quality Control
4. Marine Fisheries Research
5. Impact of Fisheries on Ecosystem Study - Marine Protected Areas (MPA) Efficiency
6. Maritime Surveillance (MCS)/Safety at Sea
7. Marine Fisheries Policy Making/Management
8. Marine Aquaculture Research
9. Marine Environmental Research

The stakeholders active in the above domains can be grouped differently to highlight how data is used. When looking at the purpose for which the data is used, there are three main categories:

- Monitoring organizations: national public sector
- Management organizations: (a) provide fisheries management recommendations / vote, (b) provide plans/measures: national and regional governmental organizations
- Exploitation companies: mainly private sectors - individual, SME and industries

Public entities are those that monitor and manage data:

1. Fisheries management institutions/Authorities
2. Fisheries Research Institutes
3. Regional Fisheries Management Organizations
4. Monitoring Control and Surveillance Organizations
5. National Bureau of Statistics
6. Regional organizations
7. Regional bureau of statistics
8. International Organizations

Private entities are those that exploit the data:

1. Individuals
2. SMEs and industries

Therefore, we can observe that there are eight different entity (public) types that monitor and manage data, while there are two entity (private) types that exploit data.

The following section will zoom in on the type of data the stakeholders work with and the relationship they have with the data, whether they are owners, processors or consumers.

5.2 Data Dimension

The data worked with can be categorized in three types:

1. Detailed data (raw or individual): data coming from an identified source (individual or vessel or company)
2. Aggregated data / statistics: data processed from detailed data
3. Reference data, standard or classification: reference data to collect and process statistics in a harmonized and standard way (list of species, types of companies, fishing areas)

Below is an outline of the relationship (owner, processor, and consumer) the stakeholders have with the data, and the type of data they work with (detailed, aggregated data/statistics, reference data, standard or classification.)

5.2.1 Data Owners

- Detailed (raw or individual) data: VMS³⁰, eRS³¹, Landing, catches, observer data
 - Individual vessels - the main data owner (provide catch, landing, positioning – all highly confidential data)
 - Fisheries exploitation related private companies - fishing vessel building, hardware companies, software companies
 - Port Authorities - authorized fishing vessels to sail
 - National Maritime/fisheries Authorities - authorized fishing vessels to fish
 - Research institutes - scientific data owner (for stock assessment)
- Statistics / aggregated data
 - International Organization - own standards
 - National Fisheries Authorities - official fisheries statistics
 - National Bureau of Statistics - national official fisheries statistics (for national account)
 - Regional Statistical Organizations - Regional statistics
 - Research Institutes - national stock assessment recommendations
 - Regional Fisheries Management Organizations - Regional statistics (in their competence area, limited number of stocks)

³⁰ VMS (Vessel Monitoring System) is a satellite-based monitoring system which at regular intervals provides data to the fisheries authorities on the location, course and speed of vessels

³¹ Electronic reporting system

5.2.2 Data Processers

- National Fisheries Authorities - fisheries statistics for national fisheries management
- National Bureau of Statistics - official statistics / national account
- Regional Statistical Organizations - official Regional statistics
- Regional Organization - collate and aggregate data for third party
- Research Institutes - stock assessment
- Regional Fisheries Management Organizations - stock assessment
- Software companies - can also provide data processing capacities such as VMS

5.2.3 Data Consumers

- National Fisheries Authorities - consume data to produce fisheries statistics for fisheries management and to supply statistics to National Bureau of Statistics
- National Bureau of Statistics - consume either detailed or aggregated data for official national statistics publication (depends on their mandate) and national account
- Regional Statistical Organizations - official Regional statistics
- Research Institutes - consume scientific data for national stock assessment
- Regional Fisheries Management Organizations - consume scientific data for regional stock assessment
- Regional Organizations - consume international standards for recommendation to member states
- International Organizations - consume national statistics to feed international databases
- Software companies - consume standards from reference sources

5.3 Stakeholder Characterization

To add another dimension to the characterization of the stakeholders, below is a list of the ten entity types with a description of (a) the activity carried out with the data used, (b) the scope of data, and (c) the purpose.

| Monitoring and Management Stakeholders | |
|---|---|
| Fisheries management institutions/Authorities | <ul style="list-style-type: none"> ▪ Activity - deliver fishing licenses, define fisheries management measures, provide recommendation for regulation ▪ Scope of data - national ▪ Why - national sustainable fishery management ▪ i.e. DPMA³² |

³² Directory for Marine fisheries and aquaculture

| Monitoring and Management Stakeholders | |
|---|--|
| Fisheries Research Institute | <ul style="list-style-type: none"> ▪ Activity - monitor fisheries, provide recommendations for fisheries management ▪ Scope of data - national data ▪ Why - assess resources (regional/national/global) ▪ i.e. IFREMER³³, IRD³⁴, INFRA |
| Regional Fisheries Management Organization (RFMO) | <ul style="list-style-type: none"> ▪ Activity - provide recommendations on stocks (stock assessment) ▪ Scope of data - regional data ▪ Why - regional sustainable fishery management ▪ i.e. IOTC/ICES³⁵ |
| Monitoring Control and Surveillance Organization | <ul style="list-style-type: none"> ▪ Activity - monitor and control fisheries activities in the EEZ countries from entities mandated from government ▪ Scope of data – regional/national ▪ Why - to enforce compliance of fishery activity with regulations ▪ i.e. NATO³⁶ |
| National Bureau of Statistics | <ul style="list-style-type: none"> ▪ Activity - centralize or produce official statistics from fisheries institutions data/statistics ▪ Scope of data - national ▪ Why - to produce evidence-based statistics for policy making ▪ i.e. INSEE³⁷, ISTAT³⁸ |
| Regional Organization | <ul style="list-style-type: none"> ▪ Activity - centralize data through definition of data calls and standards, which are shared with state members, fishery management (harmonization) for EU level with a global impact ▪ Scope of data - EU and global view (fleets operating at globally) ▪ Why - policy making at regional level ▪ i.e. EC DG-MARE³⁹, JRC⁴⁰ |

³³ <http://wwz.ifremer.fr/>

³⁴ <https://www.ird.fr/>

³⁵ <http://www.ices.dk/Pages/default.aspx>

³⁶ North Atlantic Treaty Organization <http://www.nato.int/>

³⁷ <http://www.insee.fr/fr/>

³⁸ <http://www.istat.it/en/>

³⁹ http://ec.europa.eu/dgs/maritimeaffairs_fisheries/index_en.htm

⁴⁰ <https://ec.europa.eu/jrc/>

| Monitoring and Management Stakeholders | |
|--|---|
| Regional bureau of statistics | <ul style="list-style-type: none"> ▪ Activity - centralize statistics from national level (no production of data) ▪ Scope of data- national ▪ Why - produce evidence-based stats for policy making ▪ i.e. EUROSTAT⁴¹ |
| International Organizations | <ul style="list-style-type: none"> ▪ Activity - centralize national statistics/ provide international classification ▪ Scope of data –global view ▪ Why - sustainable fishery management ▪ i.e. FAO, NATO |
| Exploitation Stakeholders | |
| <p>The entity types that exploit the data are Individuals (single persons), SMEs and Industries. The type of data exploited includes: catch, effort, landing, VMS, eRS, MCS operation. Below are two examples of how the data is exploited:</p> <ol style="list-style-type: none"> 1. Fishery - a vessel owner (industry) needs software and capacity to collect and centralize VMS data, an SME sells the software and processing capacity to do this, the company buys the software. 2. Aquaculture - monitoring and managing is mainly public, however, production is mainly private. A typical new farm needs to get a pollution production assessment study to comply with regulation; it buys it from a private consultancy company. | |

5.4 Stakeholder Mapping- Data Value Chain

To give an overall view of the two main sub-sectors (Fisheries and Aquaculture) of the Fishery and Marine Sciences Data Analysis Sector, below are two diagrams.

The arrows in the following two figures indicate the flow of data/information, beginning with raw/detailed data to an aggregated form to indicators then policies.

5.4.1 Fisheries Data Value Chain

⁴¹ <http://ec.europa.eu/eurostat>

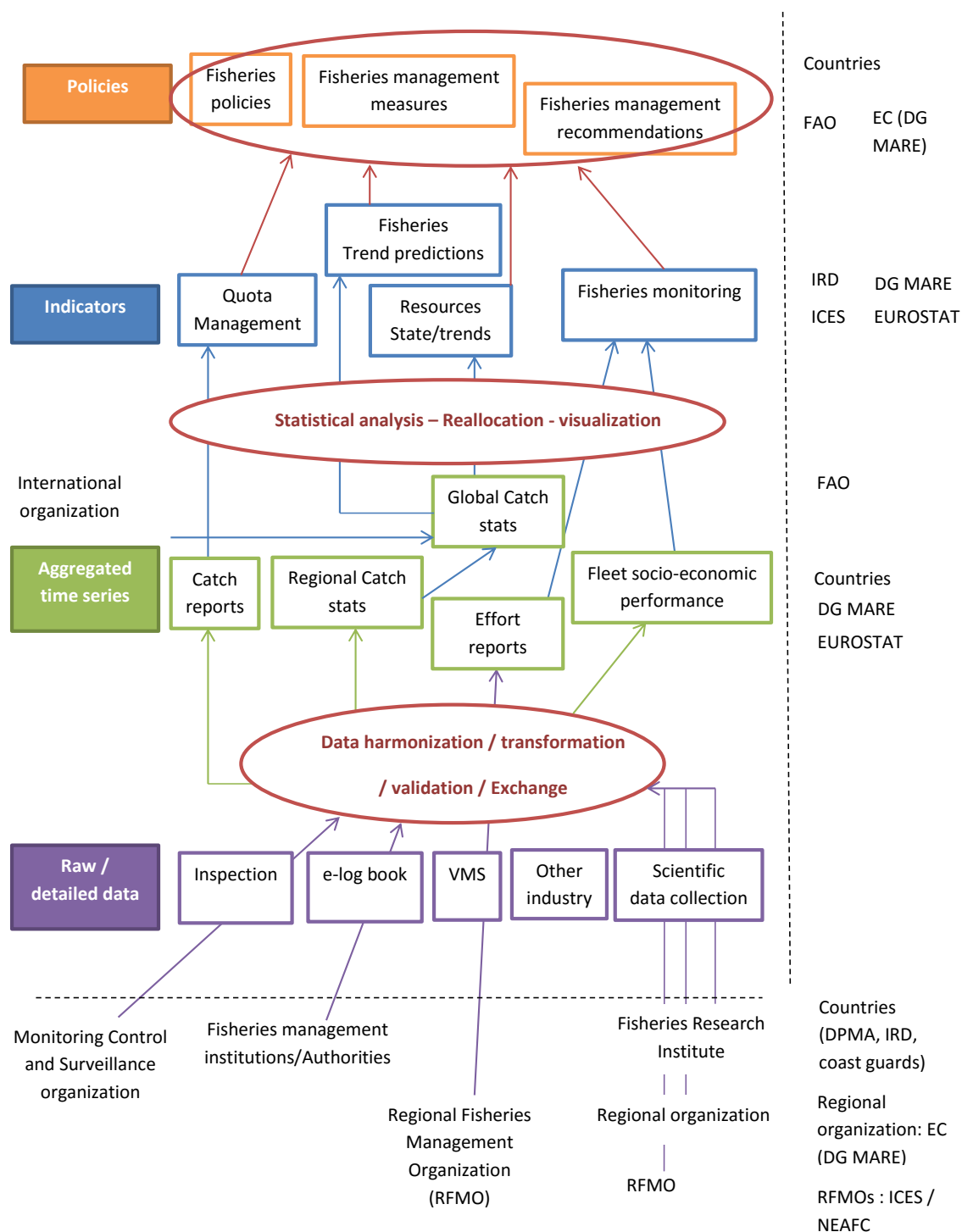


Figure 1 - Fisheries Data Value Chain

5.4.2 Aquaculture Data Value Chain

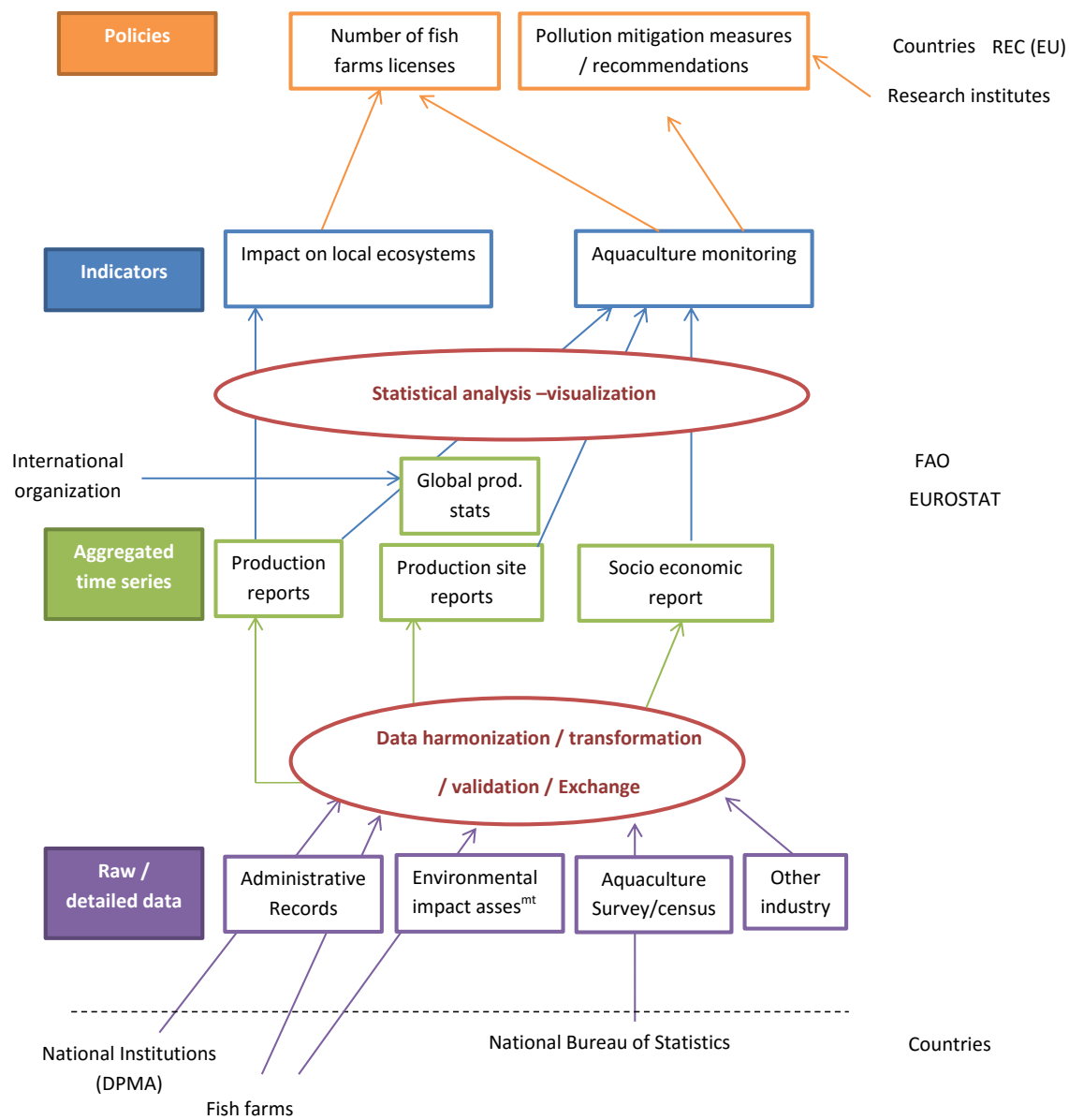


Figure 2 - Aquaculture Data Value Chain

5.5 Data Flows

This section focuses on the flow of data and information (from collection to exploitation) between stakeholders identified in this chapter for three selected domains.

For each figure there is a vertical correlation with the flow and the level of aggregation of data (bottom: detailed data - up: highly aggregated data). The figure's verticality represents also its temporality: data are collected from the bottom on a given frequency (mostly on a daily/weekly basis), processed by the intermediate level later in time and less frequently (monthly basis) and finally collated at international level on a yearly basis.

Each figure is accompanied by a thorough description. In exploring the flow of data in the given domain, a clearer picture is given and initial findings are described.

5.5.1 Fisheries Exploitation and Monitoring Domain

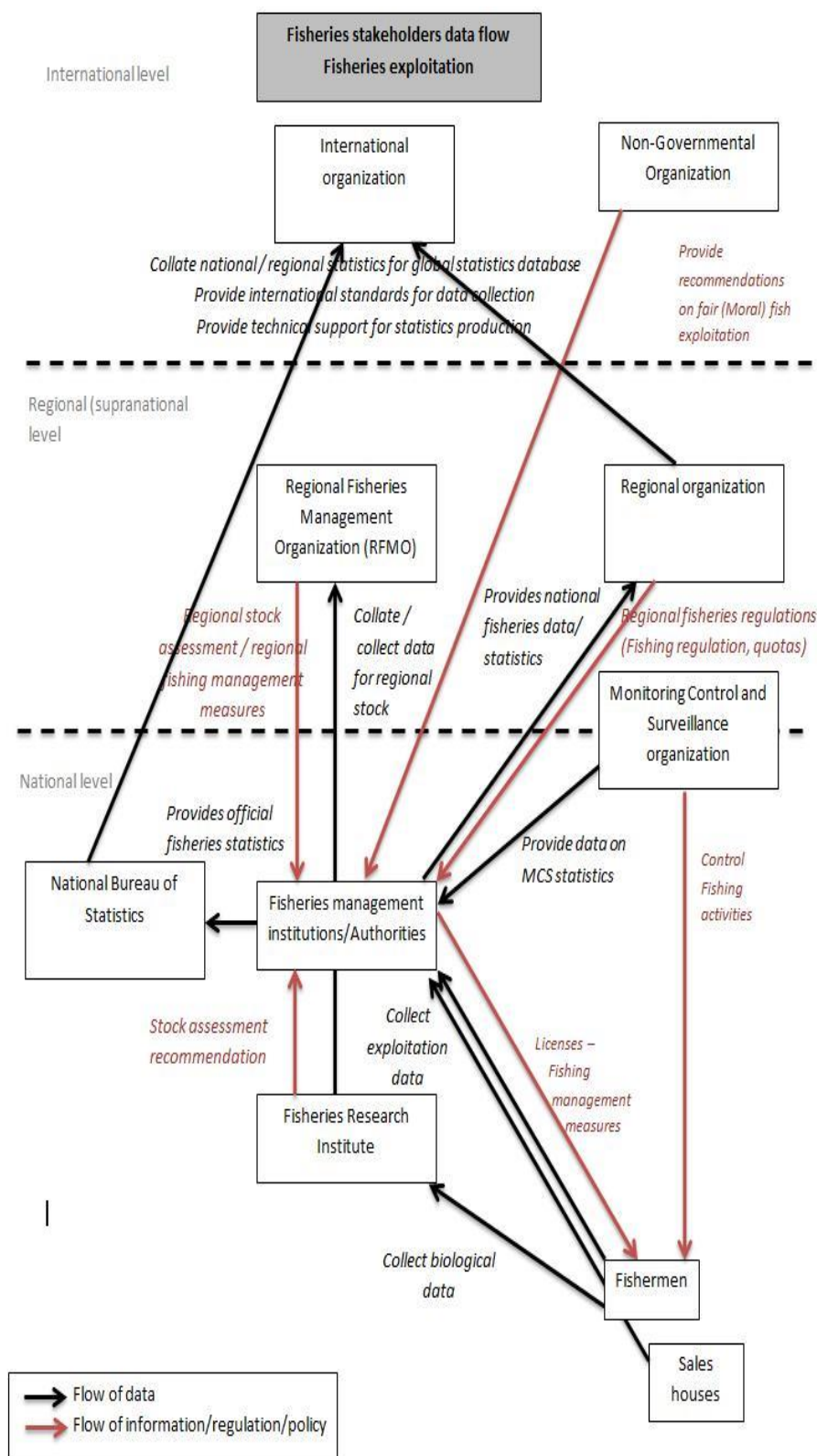


Figure 3- Fisheries Exploitation and Monitoring Domain- Data Flow

The Fisheries Exploitation and Monitoring domain has to conduct different data-related activities:

- **Assess the exploited stock:** research institutes collect scientific data to feed stock population assessment processing. They define complex population model, run the model with existing/known data for exploited stock. The result is a picture of the exploited stock (total population/school, level of reproduction, natural mortality, and acceptable level of fish taken by fishing activities). Indicators such as maximum sustainable yield are published.
- **Assess the level of exploitation of the exploited stock:** national institutions collect data from different sources (small-scale fisheries, industrial fisheries) with different methodologies / sources (log book, sample based survey, e-log book, VMS) to feed exploitation indicators computation (effort, catches, Catch per Unit of Effort). The result is the quantity of fish taken from the exploited stock. These indicators are made available to data consumers (research institutes) through formal or informal processes and some are published.
- **Assess the level of sustainable exploitation of the exploited stock:** research institutes, national institutions and regional organizations collect scientific data on exploitation (length distribution), collate data from different sources to compute stock assessment indicators (stock status) and provide recommendations for stock conservation measures (Management plans, Fisheries management measures, setting Marine Protected Areas, etc.).
- **Assess the level of control of the exploitation of the exploited stock:** national institutions enforce national and regional fisheries management measures and ensure that the stock is sustainably exploited. Data from these control operations are collected by National Institutions to monitor fisheries control activities by computing simple indicators (number of control operations, number of surveillance units, number of fishing vessels controls, in infraction, fined, etc.).

At a national institution level, most of the tools developed for data collection/processing/storage/dissemination are homemade (by CERIT in France for instance, a government IT department for Ministry of Agriculture or by IFREMER⁴²) or outsourced. A key dimension to be considered here: data confidentiality, especially fisherman catches declaration. This confidentiality is certified by governmental institutions/law and is a strong pre-requisite to any data declaration by fishermen.

The main challenge is data collection: human and financial resources needed are high given the diversity of data needed to feed a comprehensive national fisheries monitoring system. There is a need to store large amounts of data, mostly confidential. Processing and computing needs vary from one institution to another, mainly coming from by Research Institutes. Collation of data from different sources and its analysis to provide recommendations of stock assessment and stock management (policies) is a challenge for developing countries. Europe is providing an answer with the new Integrated Fisheries Data Management Programme (ex-FLUX).

⁴² <http://wwz.ifremer.fr/>

5.5.2 Fisheries Catches: Traceability/Certification/Quality Control Domain

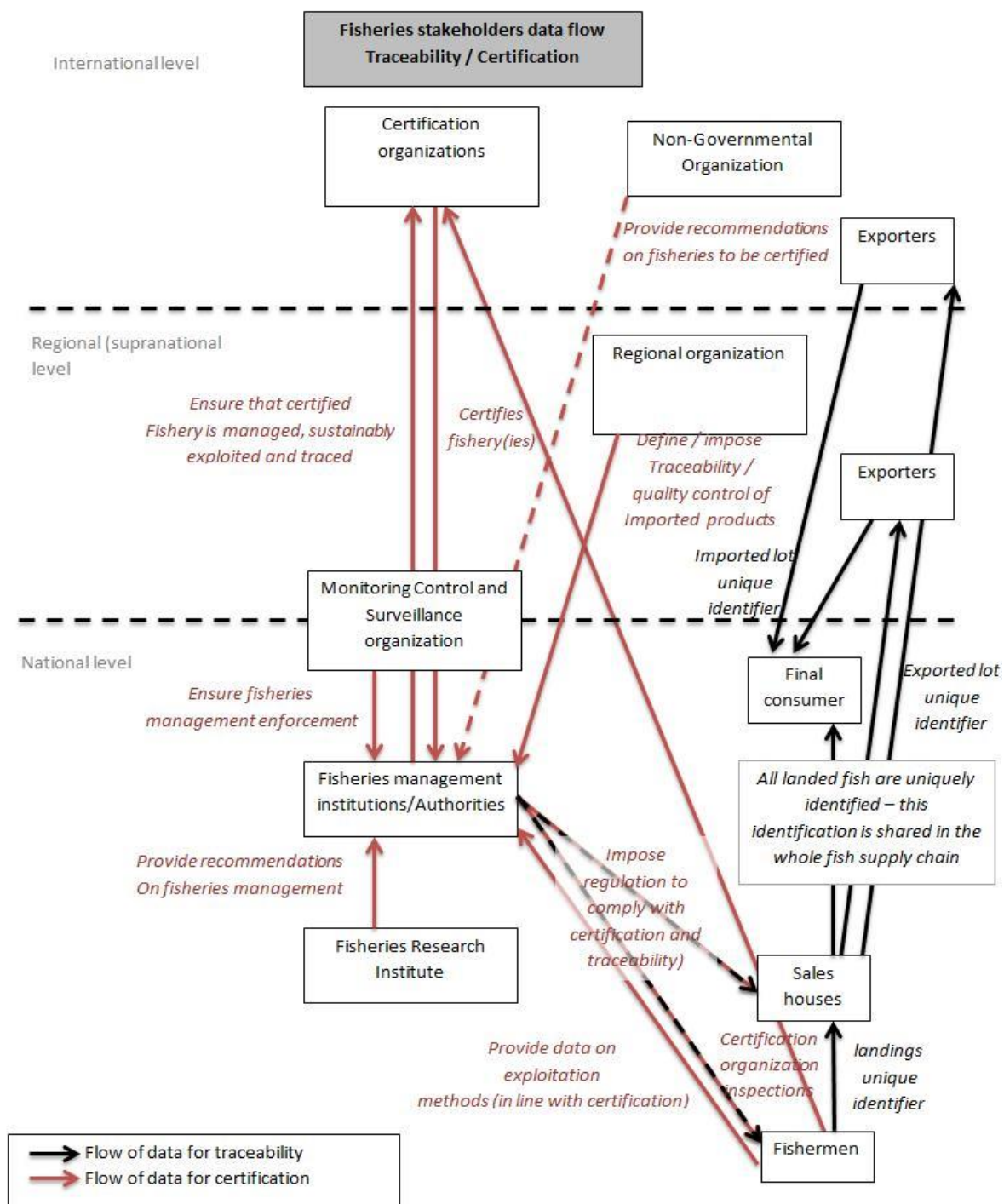


Figure 4- Fisheries Catches: Traceability/Certification/Quality Control- Data Flow

Certification aims to provide the final consumer of fish with assurance of the quality of the fish and its sustainable exploitation.

Activities are conducted at two different levels:

- **At an institutional level, certification collates information and data on the sustainability of the exploitation stock:** is it well defined, is it monitored, is it managed, is it scientifically assessed, what is the stock status? etc. This collation is highly manual (Auditor is sent to the country by the certification organization). Resulting data are stored and made publicly available to validate certification.
- **At an exploitation level, certification aims to trace each landing to ensure that every sold fish under the certification scheme is traced back to the certified stock:** when caught, when sold to the sale house or the fish monger, when sold to the fish monger by the sale house and when finally sold to the final consumer - a traceability system is in place to define a unique identifier for each landing which should be kept attached to each lot made from the initial landing and sold to and by different intermediate buyers. Each identifier must be stored and kept attached to every fish lot at each step before being made available to the final consumer.

This domain needs to store various types of data for a long term, with facilities to update the data/information easily. The certification processing is highly manual and several software suites exist for traceability.

5.5.3 Marine Fisheries Research Domain

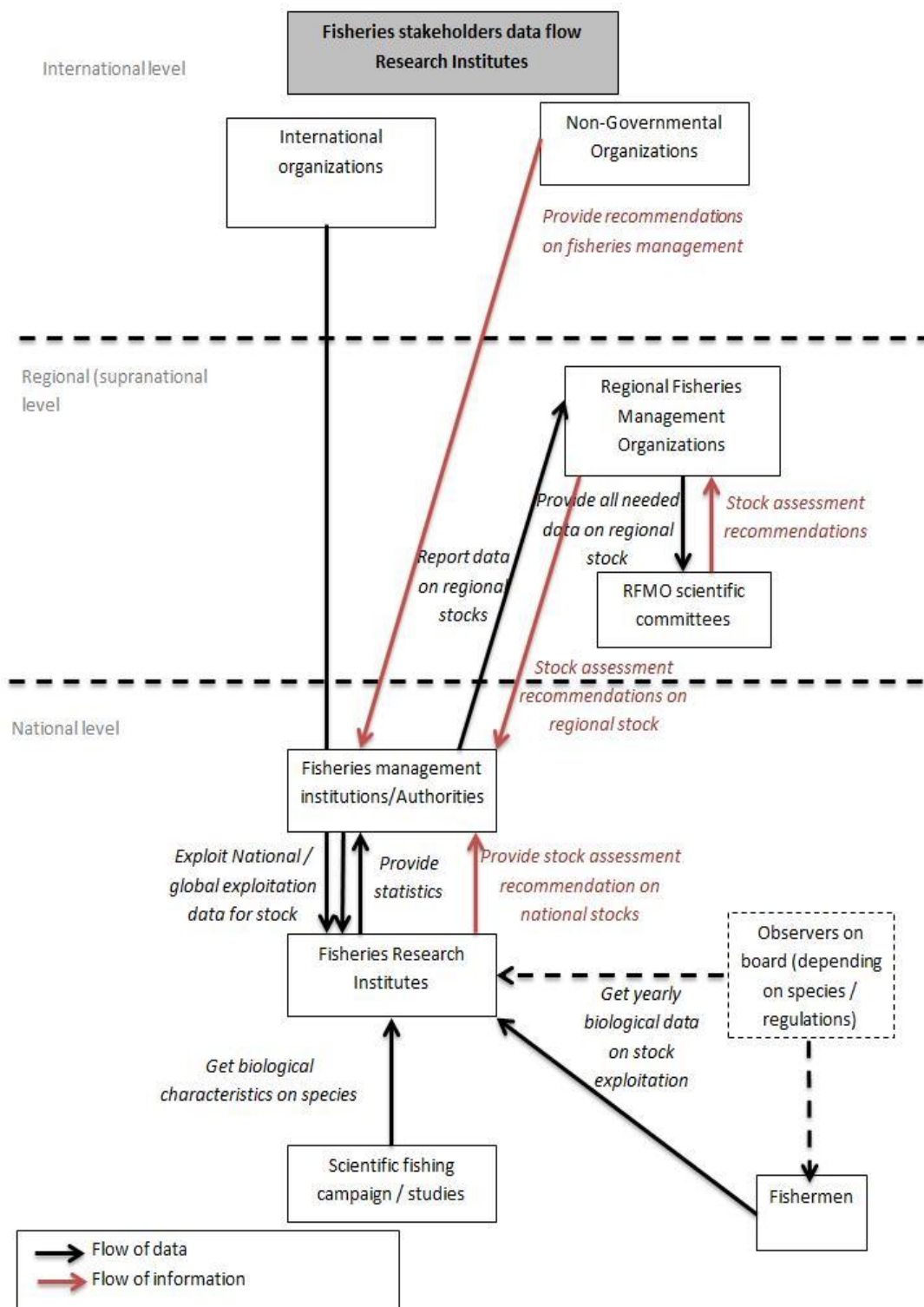


Figure 5- Marine Fisheries Research Domain Data Flow

Marine Fisheries Research activities have different scopes:

- Fundamental research on taxonomy, on fish biology/genetics, on fish population modelling
- Practical research on fishing techniques, on fishing vessels, fish mortality per fishing techniques, fish catch monitoring (developing statistical models depending on fishing type), and stock assessment

Fundamental research usually feeds practical research. Research is built on a common scheme of setting a problem, identifying experiment to collect raw data on the problem, defining model to exploit data and compute statistics with the collected data to provide assessment of the problem solution. High volume of raw data is collected, important computing resources could be needed (population model computing for instance) and limited numbers of statistics are produced.

In Marine Research Institutes tools and processing capacities needs vary from one research department to another. Mostly, ad-hoc solutions are developed under regular programmes or projects funds. There is a common use of open-source solutions (R for statistical processing, postgresSQL⁴³ for DBMS). Outsourcing of IT activities could be a source of revenue.

Research is mostly conducted using “homemade” tools (using R for statistical processing) and local infrastructure. In certain cases (like IRD in France), IT is now not considered to be part of the research institute core competencies and is outsourced.

5.6 Domain and Stakeholder Findings Representing Opportunities for EGI

Thus far, through the analysis of domains in the Fishery and Marine Sciences Data Analysis Sector and its Stakeholders we have discovered that Fishery Exploitation is mainly, if not exclusively, private (individual, SME and industry), whereas Monitoring and Management is mainly public (national and regional governmental/public organizations).

EGI might look to focus on high interest domains that exist under Monitoring and Managing. The five domains below are where a large volume of data is applicable:

1. Marine Fisheries Exploitation and Monitoring, especially for industrial fisheries
2. Marine Fisheries Research to provide stock assessment, where there are few types of data but with large volume
3. Fisheries Catches Traceability/Certification/Quality control, where there are a large number of different types of data from different stakeholders
4. Marine Environmental research, where there is a need for harmonized data by public companies
5. Marine Aquaculture Research

⁴³ <http://www.postgresql.org/>

In addition, the following domains present conditions where EGI may find opportunities:

1. Marine Aquaculture Production and Monitoring, as main data providers in aquaculture with different challenges in terms of data storage, processing, dissemination and innovation for data monitoring
2. Impact of Fisheries on Ecosystem Study – Marine Protected Areas (MPA) Efficiency, as a large quantity of ad-hoc data is collected
3. Maritime Surveillance (MCS)/Safety at Sea, as a current challenge is processing data processing coming from different sources
4. Marine Fisheries Policy Making/Management, as the EC is in the process of standardizing fisheries reporting formats from all stakeholders (Member States, RFMO, etc.) through UN/CEFACT⁴⁴ standards and this could create business

Through the analysis of the data dimensions we understand that the scope of data worked with is often also global, this gives us insight into what services EGI may be able to offer in the future.

The mapping of the Fisheries and Aquaculture data value chain, in addition to the Data Flows created in the following three domains, gives us insights into challenges, current progress, and possible opportunities:

1. **Marine Fisheries Exploitation and Monitoring Domain presents a main challenge, being data collection:** There is a need to storing large amounts of data, mostly confidential. Processing and computing needs vary from one institution to another, mainly coming from Research Institutes. Europe is providing an answer with the new Integrated Fisheries Data Management Programme (ex-FLUX).
2. **Fisheries Catches: Traceability/Certification/Quality Control Domain** requires a need to store various types of data for long term duration. Also, facilities are needed to update this data/information easily. The certification processing is highly manual. In regards to traceability, several software suites exist.
3. **Marine Fisheries Research Domain present typical needs of research:** capacity to store large amounts of raw data for long term duration, capacity of computing models and capacity to share/exchange data with colleagues.

In Marine Research Institutes tools and processing capacities needs vary from one research department to another. Mostly, ad-hoc solutions are developed under regular programmes or projects funds. There is a common use of open-source solutions (R⁴⁵ for statistical processing, postgresSQL for DBMS). Outsourcing of IT activities could be a source of revenue.

The use of external infrastructure for cost efficiency could be promoted. Revenues will be generated by providing these institutions support to store and process their data, possibly to access more advanced pieces of software, and long term maintenance.

Access to external data is also a need, but most of external data used by these institutions are free (FAO data), it will be difficult to ask them to pay for something they can have for free.

⁴⁴ <http://www.unece.org/cefact>

⁴⁵ <https://www.r-project.org/>

6 Survey Analysis

A bottom-up approach was taken here by having direct contact with the stakeholders. This Chapter presents the results of the questionnaire circulated and the interviews performed. Furthermore, a case (project) which serves the Fishery and Marine Sciences Data Analysis Sector is presented.

6.1 Approach and Objectives

The scope of the survey is limited to data managers and owners in the Fishery and Marine Sciences Data Analysis Sector who ideally have enough understanding of EGI related infrastructures and are involved in technical decisions. This approach led us to carefully choose our respondents, concentrating more on quality, not quantity.

All nine domains identified previously as potential high interest to EGI are represented in the stakeholders who replied to the questionnaire.

The objective of the questionnaire is to understand the computing needs and obstacles of our stakeholders to gain insights and understand how EGI could better serve the community.

In the definition of the method, question content, and question wording, the following guidelines were taken into account:

- Collect exploratory (qualitative) information through some open-ended questions (short and neutral);
- Questions must be: fully understandable by the respondent, not ambiguous, and truly needed;
- Encourage respondents to provide accurate and complete info, also by putting the questions into a meaningful psychological order.

6.2 Analysis

The analysis consists of comments on each individual question, and an overall analysis (findings), which includes reflections of the interviews, this presented directly in Chapter 8.

6.2.1 Questionnaire

Below is a quick overview of the questionnaire⁴⁶ followed by an analysis of all the answers, minus those strictly related to the individual, such as Name and Role held in the organization.

Questionnaire Structure

⁴⁶ Link to the online questionnaire <https://goo.gl/i26neK>

- 26 questions grouped in four sections:
 - Section A: General
 - Section B: Data Management Aspects
 - Section C: Cloud-Based Services
 - Section D: Your Cloud Needs
- To ensure a qualitative approach, almost half of the questions are either completely open (free text), or have a free text field, being “other”.

Respondents

- Of the 21 questionnaires sent, we received 14 replies from three different sectors: 2 private non-for profit, 2 private profit, and 10 public:
 1. GRID-Arendal⁴⁷ – private non-for-profit
 2. I2S⁴⁸ – private profit
 3. IOTC⁴⁹ (Indian Ocean Tuna Commission) – public
 4. IRD⁵⁰ (Institut de Recherche pour le Développement) – public. Note: Two separate replies.
 5. CLS⁵¹ (Collective Localisation Satellites)– private profit
 6. MSC⁵² (Marine Stewardship Council) – private non-for-profit
 7. HCMR⁵³ (Hellenic Centre of Marine Research) – public. Note: Two separate replies.
 8. JRC⁵⁴ (Joint Research Centre - European Commission) – public
 9. DPMA/MSIPA⁵⁵ (French Ministry of Ecology, Sustainable Development and Energy)– public
 10. Ecopath International Initiative⁵⁶ – public
 11. European Commission – public
 12. IAEA⁵⁷ (International Atomic Energy Agency) – public

Infrastructure Awareness When asked in Question 7 - *How familiar are you with the existence and service catalogue of the following three e-Infrastructures? EGI, PRACE⁵⁸, and GEANT⁵⁹:*

- 8 respondents have never heard of any of the infrastructures
- 4 respondents either heard of, investigated or have a vague idea of GEANT
- 1 respondent heard of EGI
- 1 respondent heard of PRACE

⁴⁷ <http://www.grida.no/>

⁴⁸ <http://www.i2s.gr/>

⁴⁹ <http://www.iotc.org/>

⁵⁰ <https://www.ird.fr/>

⁵¹ <http://www.cls.fr/en/>

⁵² https://www.msc.org/it?set_language=it

⁵³ <http://www.hcmr.gr/gr/listview2.php?id=8>

⁵⁴ <https://ec.europa.eu/jrc/>

⁵⁵ <http://www.developpement-durable.gouv.fr/>

⁵⁶ <http://www.ecopath.org/consortium>

⁵⁷ <https://www.iaea.org/>

⁵⁸ <http://www.prace-ri.eu/>

⁵⁹ <http://www.geant.net/>

6.2.1.1 Section B- Data Management Aspects

Question 8 - What is the current scope of the data you manage?

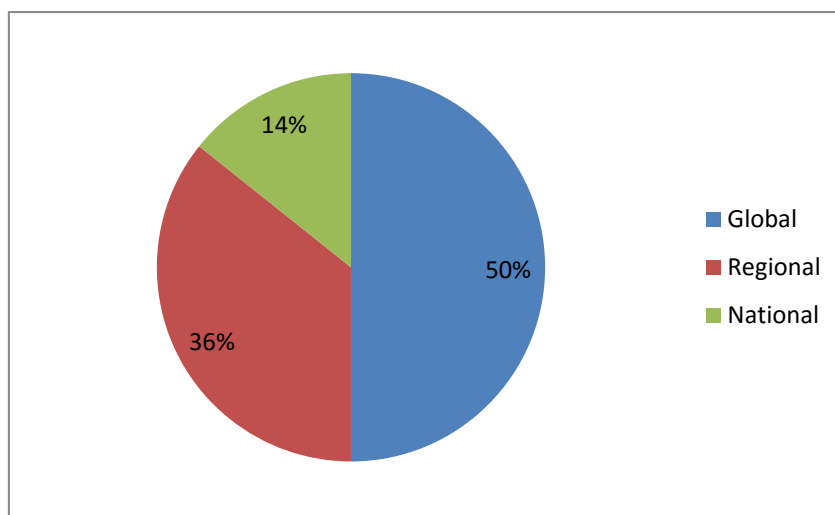


Figure 6 - Question 8 - What is the current scope of the data you manage?

Data worked with is increasingly global, at 50 percent respondents also require the right tools to manage the data. Regional (supranational-European Union) is in second rank at 36 percent.

Question 9 - What type of data does your institution/company manage? Tick all that apply.

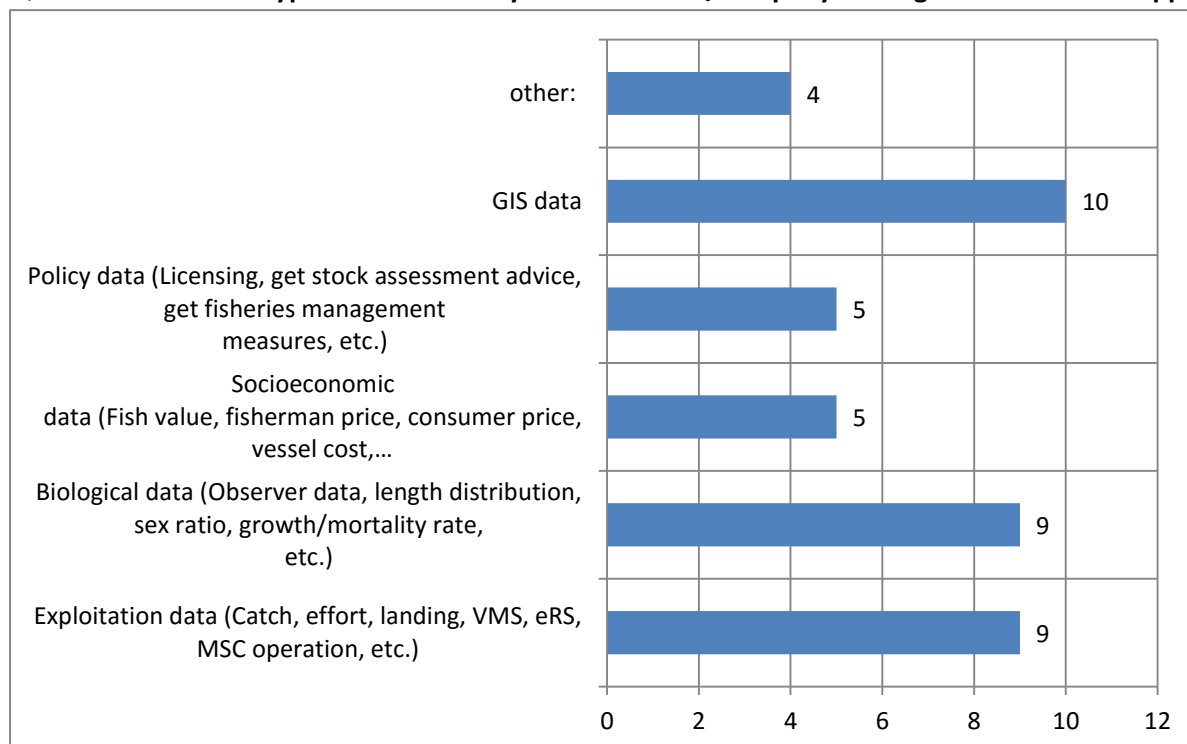


Figure 7 - Question 9 - What type of data does your institution/company manage?

GIS⁶⁰ data, at just over 71 percent, is the data most managed. In second rank, is Exploitation and Biological data at just over 64 percent. To note, Exploitation data is at the beginning of the data value chain, therefore other types of data/information depend on it.

Question 10 - Select the top three priorities for your institution/company?

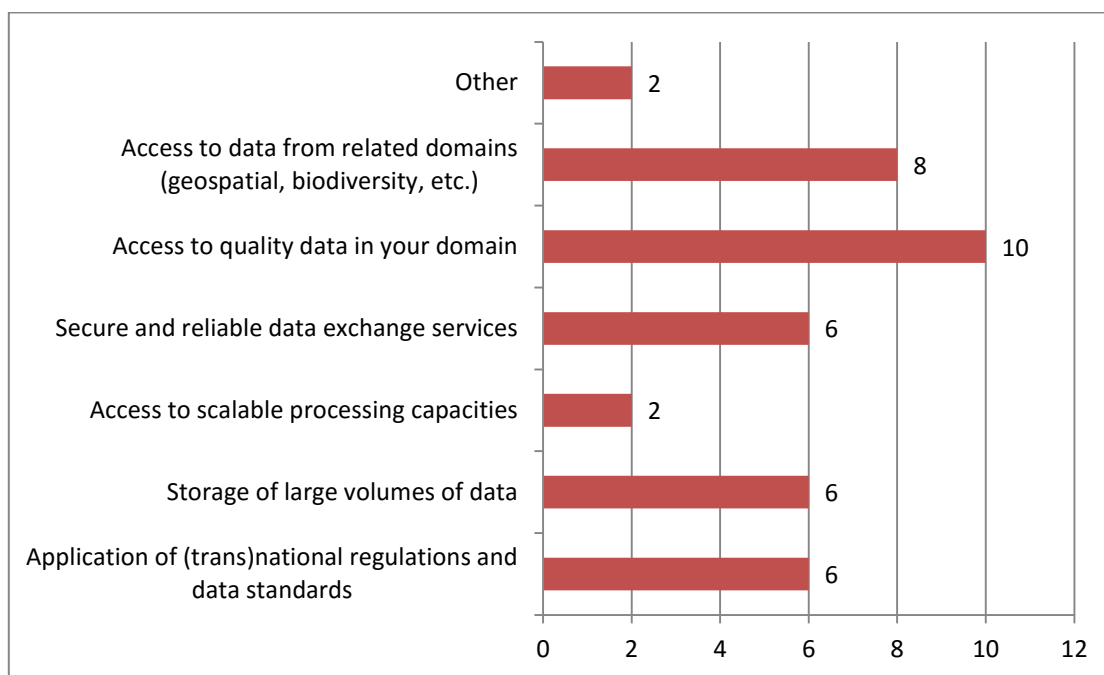


Figure 8 - Question 10 - Select the top three priorities for your institution/company?

The single top priority in data management is access to quality data in the domain, at just over 71 percent. Interestingly, in second rank, is Access to data from related domains (geospatial, biodiversity, etc.) at 57 percent. Access to scalable processing capacities was ranked last.

Question 11 - Do you need access to external data from the same domain or other domains? If yes, please list the domains, the data providers and the type of data that you are interested in.

Below are the replies of those that answered “Yes” and commented:

- Species occurrence data and environmental parameters from any provider, incl. OBIS⁶¹. Maritime traffic Remote Sensing data, NASA⁶², ESA⁶³ Taxonomic classifications and species names: WoRMS⁶⁴, Species pictures

⁶⁰ Geographic Information System, a system designed to capture, store, manipulate, analyse, manage, and present all types of spatial or geographical data.

⁶¹ <http://iobis.org/mapper/>

⁶² <https://www.nasa.gov/>

⁶³ <http://www.esa.int/ESA>

- Ecological: - species data (FishBase⁶⁵, SeaLifeBase⁶⁶, WoRMS, GBIF⁶⁷) - species distribution data (AquaMaps⁶⁸, OBIS) Environmental (marine): - primary productivity (GFDL⁶⁹, JRC) - temperatures (GFDL, others) - salinity (GFDL, others) - pH (GFDL, others) - oxygen (GFDL, others) – etc. Socio-economic: - Fishing effort (national databases, MEDITS⁷⁰) - Catches (national databases, MEDITS) - Discards (national datasets)
- FAO List of Species for Fishery Statistics Purpose (ASFIS); FAO Global Capture Production; FAO Vulnerable Marine Ecosystems Database; <http://fisheryandseabird.info/>; RAM Legacy Stock Assessment Database
- Yes from 3rd world countries
- Yes, generally any spatial data on marine regulations is very useful for our daily work. This may include fisheries regions e.g. <http://www.fao.org/geonetwork>, <http://www.marineregions.org>, <http://www.pacgeo.org>. Increasingly there is a need to be able to access the authoritative data from countries, which is generally not available.
- LifeWATCH⁷¹, IPT, Fishbase⁷², EOL, COL, BHL, OBIS, GBIF
- Data from the same domain since my work focuses on highly migratory species (i.e. tunas) that are caught by several countries all over the world - Data on price of fish on the international trade market - Bathymetric data - Oceanographic data (i.e. outputs from physical, biogeochemical and ecosystem models, remote-sensing processed data, in-situ data from oceanographic drifters for instance)
- Satellite images Model Outputs Acoustic Data Fisheries and related biological data
- Data related to quotas, vessels from EU, catch & vessel location data from others countries, scientific data from specialised bodies
- In situ data related to the domains of interest: maritime security, environmental surveillance, marine resource management.

All respondents need access to external data and many from other domains.

Question 12 - From the domains above specified in your answer to Question 9, please let us know what challenges you encounter in regards to specific regulations, standards, formats, protocols etc.

Here below are all the replies. Notable challenges encountered are underlined.

- Data security and integrity

⁶⁴ <http://www.marinespecies.org/>

⁶⁵ <http://www.fishbase.org/>

⁶⁶ <http://www.sealifebase.org/>

⁶⁷ <http://www.gbif.org/>

⁶⁸ <http://www.aquamaps.org/>

⁶⁹ <http://www.gfdl.noaa.gov>

⁷⁰ <http://www.sibm.it/SITO%20MEDITS/principaleprogramma.htm>

⁷¹ <http://www.lifewatch.eu/>

⁷² <http://www.fishbase.org/search.php>

- Open Geospatial Consortium (OGC⁷³) is the top priority to structure and expose data. Infrastructure for spatial information in the European community (INSPIRE⁷⁴) is just about being compliant with OGC within European territories which is not relevant for us. TDWG standards related to GBIF are also very interesting to disseminate some of our data. RDF&SPARQL are key, as well as exposing our data through OPEN DATA initiatives like data.gouv.fr. Having a data abstraction library to turn any standard data format into another one would be the best so that everybody can focus on some standards and get others in an automated way (kind of Gdal).
- Regulations (Control regulation), standards (UN/CEFACT)
- Standards: Although standards exist, colleagues entering data have difficulties to comply. As aggregators ourselves, we spent a lot of time in standardization and quality control. Also, standards cannot fit the richness of biodiversity data, and ad hoc solutions have to be integrated on the fly.
- The export of data in standard formats is a full-time job that requires extra human resources and is very specific. Even if one manages to have the resources to conduct the work in the first place (i.e. through short-term contracts), the evolution of the databases, extent of data collected, etc. make necessary the ability to manage such exports with local (permanent) staff, while such qualified personnel (i.e. compute scientist with appropriate skills) is missing from many teams and institutions; - There is currently few information transferred to the scientists who work on data (fisheries data in my case) but do not know much about international standards, regulations, etc. Although the information is certainly available somewhere, the development and implementation of such standards and protocols seem to be done apart from the scientists who work with the data. More information and training activities should be conducted so that scientists understand the importance of such regulations, etc. and how to access, use, and read the data for instance. Overall, there is a strong need to provide assistance to the research institutes so that they improve their understanding of the evolutions in data tools and formats.
- The three main challenges are: 1.) the availability of appropriate metadata and keeping this linked with the data products licencing; 2.) referencing of data (especially restrictions on data use) 3.) Access to data in useable formats (although data is increasingly being made available using OGC standards).
- Data collection, reporting, and low levels of data compliance by coastal states in the Indian Ocean, particularly related to coastal fisheries, remains one a significant challenge for the IOTC Secretariat.
- Regulations, formats, data protection, confidentiality rules
- Data accessibility, especially socio-economic data, is a big problem. Standards to find and integrate data must be further implemented, we would love to see OpenDAP⁷⁵/THREDDS⁷⁶ standards further promoted.

⁷³ <http://www.opengeospatial.org/ogc>

⁷⁴ <http://inspire.ec.europa.eu/>

⁷⁵ <http://www.opendap.org/>

⁷⁶ <http://www.unidata.ucar.edu/software/thredds/current/tds/>

- Regulations: None that we are aware of. We do not hold spatial data currently. Standards, Format and Protocols: We use whatever is appropriate for the situation at hand and have no issues with them specifically.
- INSPIRE, OGC
- Standards (OGC - WMS)

Question 13 - In 2-5 years from now, what do you believe will be your main challenge regarding the management of data? Tick all that apply.

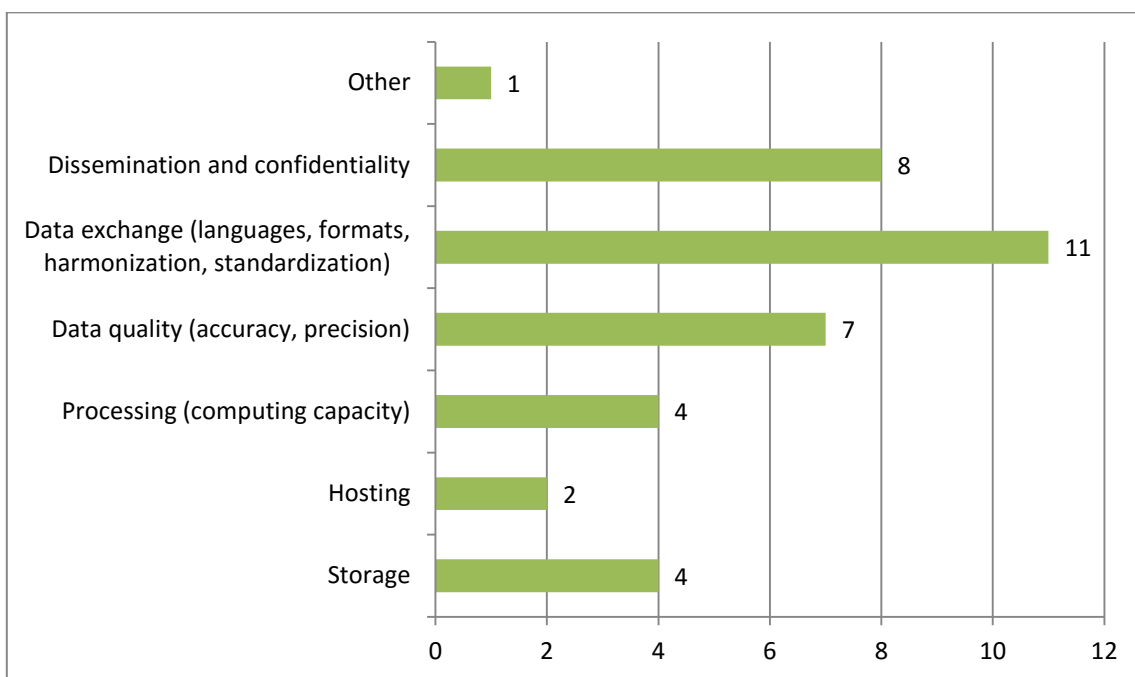


Figure 9 - Question 13 - In 2-5 years from now, what do you believe will be your main challenge regarding the management of data?

Data exchange, at almost 80 percent, remains by far the biggest challenge our respondents foresee. In second rank at 57 percent, Dissemination and confidentiality was chosen. Data quality at 50 percent is particularly important for this community to perform well. Hosting and Storage, and Processing don't worry our respondents much.

6.2.1.2 Section C – Cloud-Based Services

Question 14 - Does your institution/company use Cloud-based services?

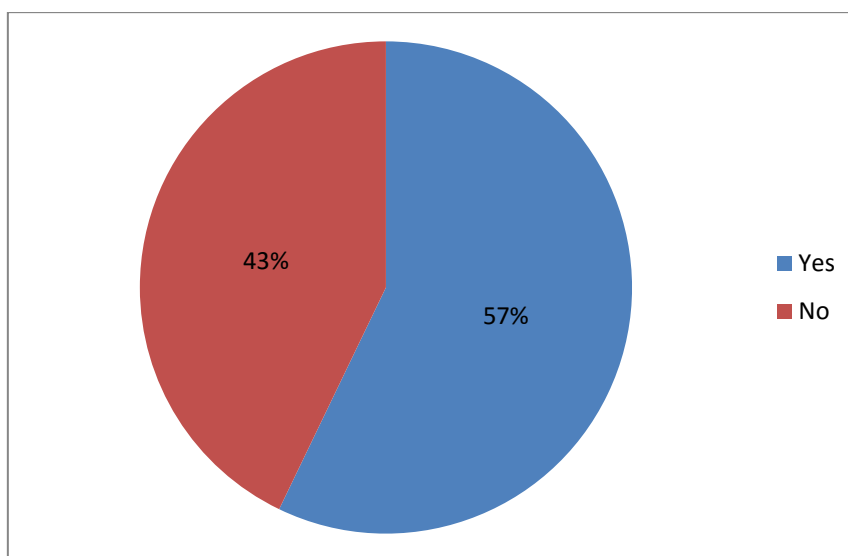


Figure 10 - Question 14 - Does your institution/company use Cloud-based services?

57 percent of respondents use Cloud-based services.

Question 15 - If your institution/company uses Cloud-based services, which services? Tick all that apply

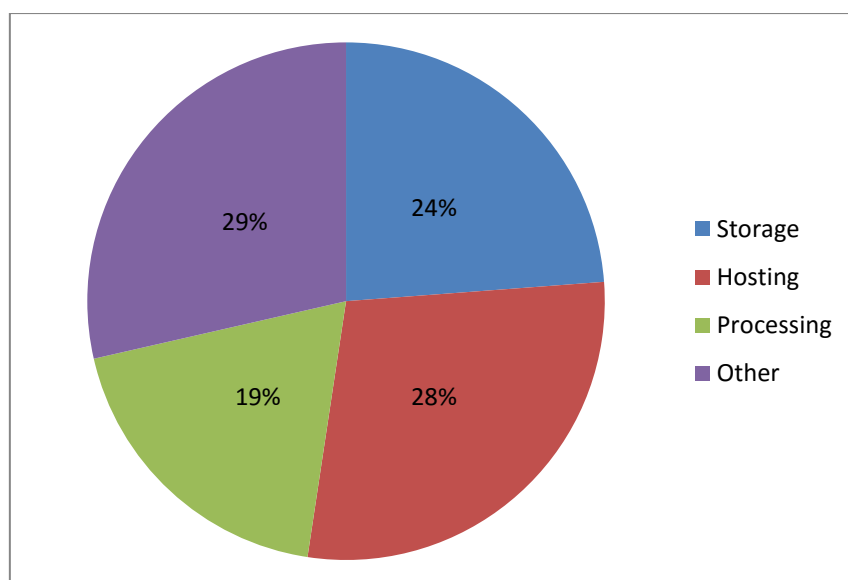


Figure 11 - Question 15 - If your institution/company uses Cloud-based services, which services?

Of those using cloud-based services (57 percent), we can observe an almost even distribution, where all services are used.

Question 16 - If your institution/company uses Cloud-based services, what type of Cloud?

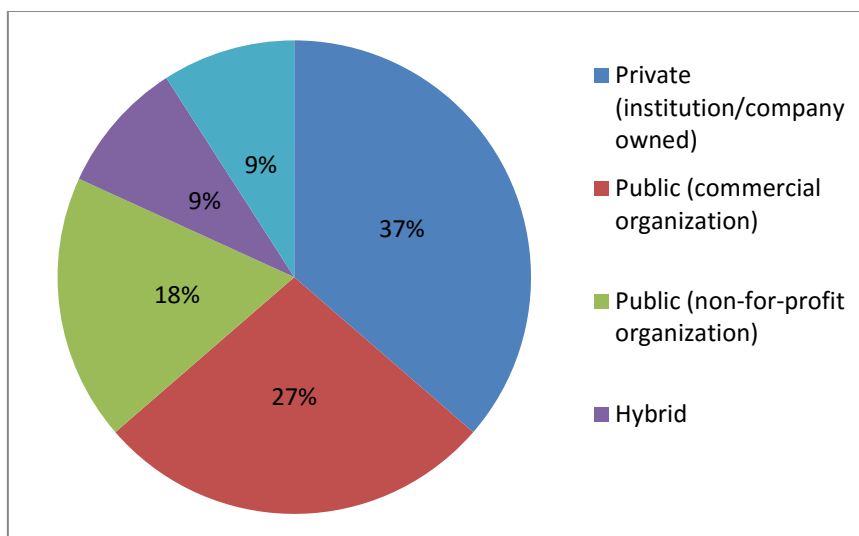


Figure 12 - Question 16 - If your institution/company uses Cloud-based services, what type of Cloud?

The majority, at 37 percent, use a private cloud. Second rank is public (commercial organization) cloud at 27 percent.

Question 17 - If your institution/company is already using some Cloud-based Services, do you plan to further invest in other services? If yes, which ones? Tick all that apply.

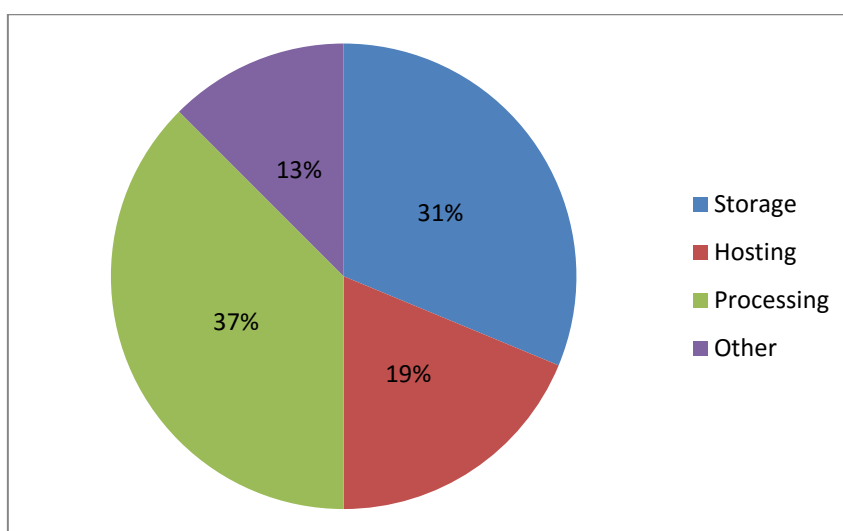


Figure 13 - Question 17 - If your institution/company is already using some Cloud-based Services, do you plan to further invest in other services? If yes, which ones?

At 37 percent, those planning to invest in other cloud-based services are looking to use processing services. 31 percent are interested in storage.

Question 18 - If your institution/company is not using Cloud-based services although is interested, what services would be of interest within the next two years? Tick all that apply.

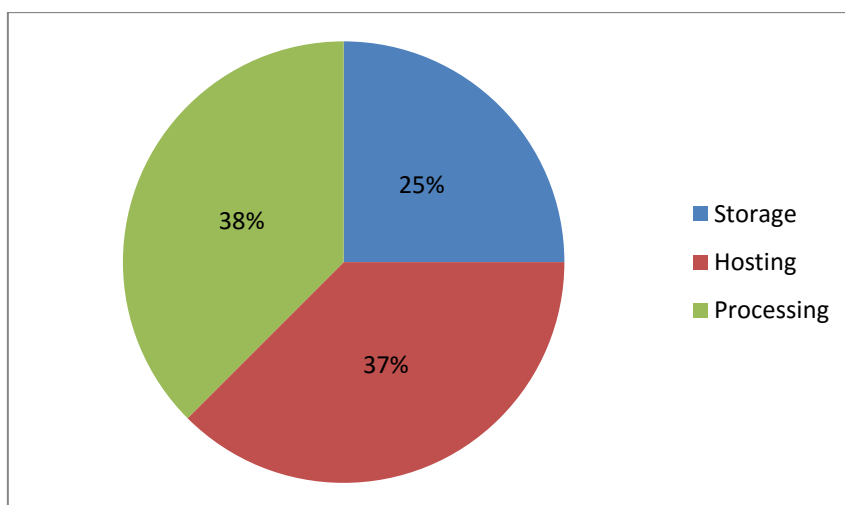


Figure 14 - Question 18 - If your institution/company is not using Cloud-based services although is interested, what services would be of interest within the next two years?

Those respondents not already using cloud-based services, but interesting in investing in the future, are leaning towards Hosting and Processing almost in equal percentage (37-38), taking priority over storage.

Question 19 - Are you aware of any obstacles/challenges your institution/company recognizes in regards to using Cloud-based services? If yes, in which areas? Tick all that apply.

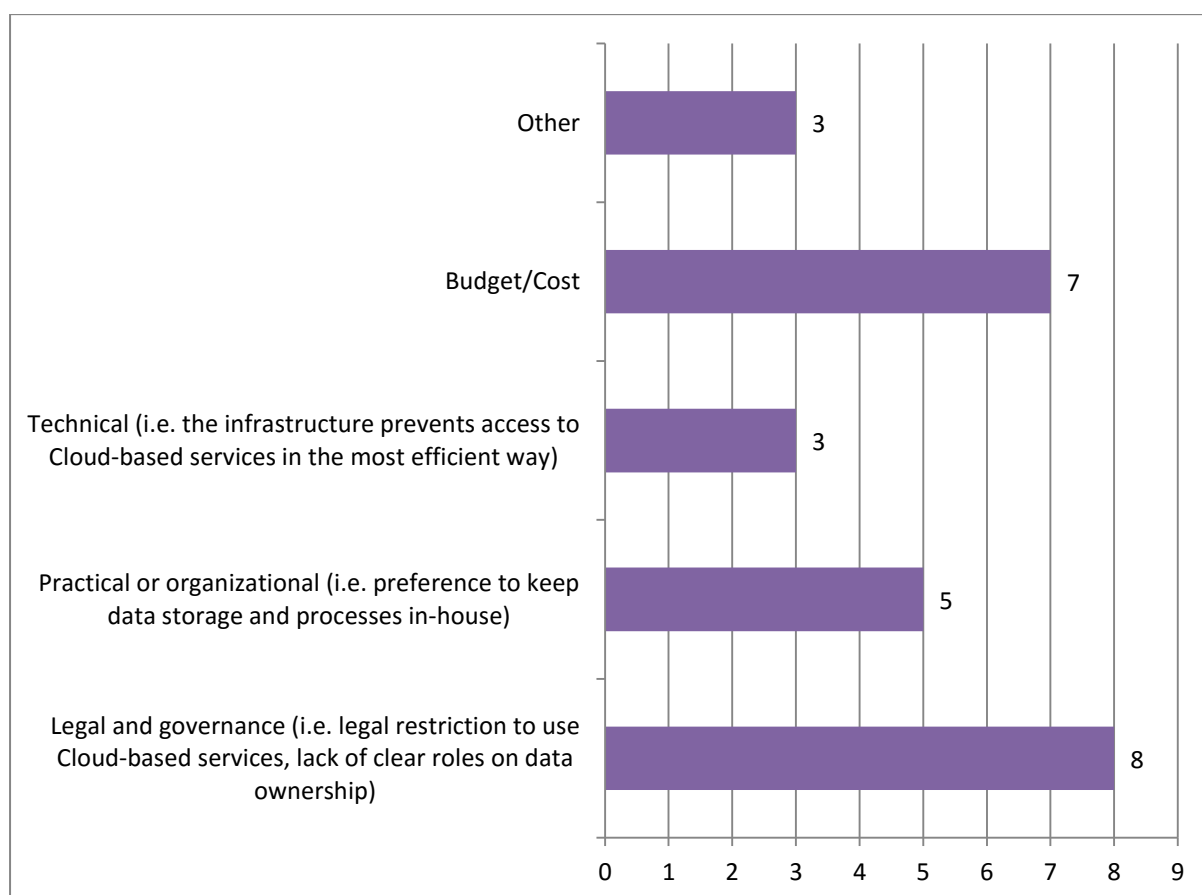


Figure 15 - Question 19 - Are you aware of any obstacles/challenges your institution/company recognizes in regards to using Cloud-based services? If yes, in which areas?

Legal and governance is the top challenges/obstacles recognized for cloud-based services at 57 percent. Budget/Cost is a natural concern, as most of the respondents represent public organizations.

Question 20 - Please provide insights about how you feel the challenges chosen in question 19 can be overcome.

Below are all the replies. The replies concentrated more about the issue(s) rather than the ideas to overcome the challenges. Underlined are the notable insights given.

- We are slowly transitioning over from self-hosting of services to utilising cloud-based services.
- Cloud fundamentally is a good idea and storage/processing, hosting is not central to many organizations’ business. Inertia exists due to legal complications/ambiguity, specific regulations in some industries that prevent, inhibit or constrain uptake and lack of understanding at some senior levels. Some companies may have invested extensively in an

on premise estate and moving such arrangement into the cloud can be complex and time consuming as many cloud solutions are limited in some way or another. This can make uptake extremely challenging even when the internal will exists. Finally, time is also required before such investments can be perceived as legacy and the business case for change made.

- A first issue would be financial, i.e. how to get the funds required to maintain such cloud systems, which need to be sustainable when most budgets of research institutions now rely on short-term project (i.e. a few years). A second issue would be to share the applications and services with other research institutes and Universities who have similar needs to maximize the interest of the approach and share techniques and processes among partners. This requires to well identify the needs of each potential partner and to implement simple technical tools that do not need high qualifications in computer science.
- We are a small not-for-profit organization and cannot afford expensive processing facilities. We are looking for affordable and scalable cloud facilities to run large ecosystem models, and even provide these services to EwE⁷⁷ (Ecopath with Ecosystem) users.
- Long-term planning
- Network of local solutions (Medium size cluster) and accessing to several clouds.
- LifeWATCH is already engaging discussion with EGI LifeWATCH Greece is also in discussion with D4Science
- The challenge to use some processing services delivered by the cloud is more in having technical advices and making the use of such service seamless for researchers.
- By cooperating with experts in the area and internal research.

6.2.1.3 Section D: Your Cloud Needs

Question 21 - On a scale from 1 to 4, how important is the Cloud Computing benefit: Facilitate analysis of cross-domain information through harmonization and standardization.

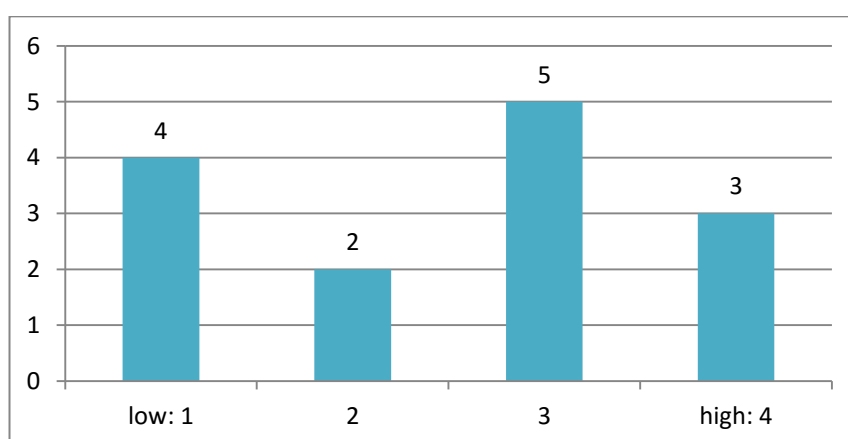


Figure 16 - Question 21 - On a scale from 1 to 4, how important is the Cloud computing benefit: Facilitate analysis of cross-domain information through harmonization and standardization.

⁷⁷ <http://www.ecopath.org/>

The respondents seem to be divided between low and high importance, although more respondents are on the higher-end.

Question 22 - On a scale from 1 to 4, how important is the Cloud Computing benefit: Cost saving

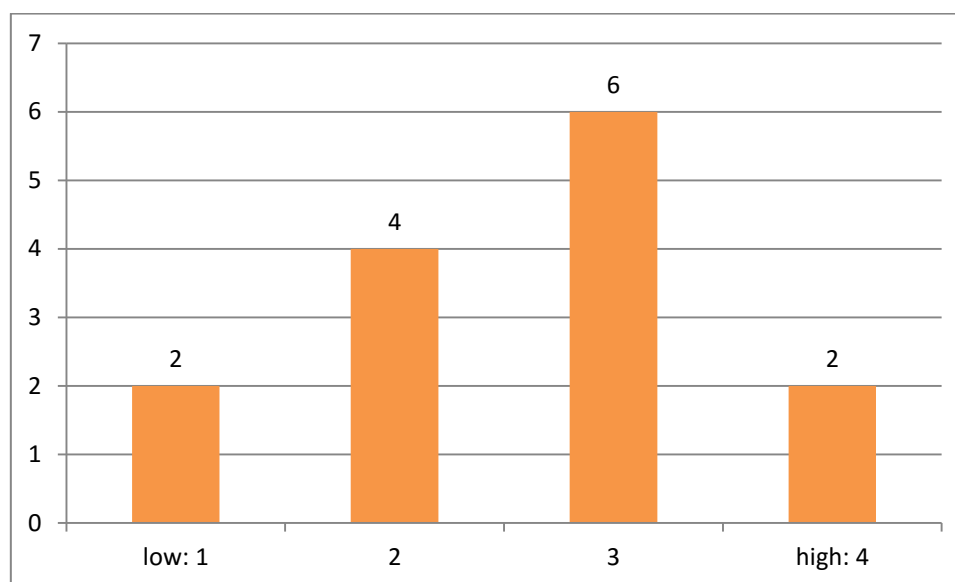


Figure 17 - Question 22 - On a scale from 1 to 4, how important is the Cloud computing benefit: Cost saving

Cost saving is quite important to our respondents, but not highly important.

Question 23 - How important is cost when choosing Cloud-based services?

Below are all the replies. Underlined are the notable replies.

- As a public research body, reducing cost has been unfortunately important for years. This is fine if we can get more services with the same budget. However, cloud solutions for Southern countries are still irrelevant in some cases due to string limitations of bandwidth or stability of Internet access.
- Very - we're poor. We cannot use services if they cost money, unless a project can pay for service access.
- Cost is important, but on the other hand there is no other alternative. It is not a feasible solution to build your own data centre, in such a way to be able to provide cloud services.
- It depends on the service; one has to balance performance, reliability and cost for every service and how crucial are these services for the related activity.
- Any cost has to be affordable and one key issue with high value cloud service offerings is they are not affordable for SMEs. Once affordability is addressed, cloud economics move organisation's from a CAPEX to OPEX cost, which is not always wanted in the not for profit/charity space. Finally, cloud providers are very distant from customer and define service in terms of "storage", etc.; when businesses need a "full service story" (i.e. advice,

consultancy, training, and help), which typically needs to be delivered by another 3rd party or upskilled in-house staff.

- The cost of long-term needs will be scrutinized in detail. The costs for one-shot or short-term needs will be less important depending on the importance or urgency of the work.
- This relates to the funding potential of the research institute itself and the choices they will make with regards to the overall needs of the researchers and administration. I would guess the cost is very important with regards to the constant reduction of budgets in public research and the IRD Direction would estimate the expected benefits for the institute in relation with the needs for other institutions at the same time.
- In my opinion, the cost should be weighted considering the impact of the results obtained (spatial and temporal valuation).
- Cost has a “high weight” when choosing Cloud-based services
- Many of the datasets collated by the IOTC Secretariat are confidential and many countries would be reluctant to report data if the storage, processing and dissemination were cloud-based rather than data stored in-house. For this reason, the IOTC Secretariat at this point has not fully explored or considered options for cloud-based services.
- Not that much
- It is important that the cost is scalable and not prohibitive. There are also challenges in using cloud services on a project basis, where there is a fixed time period of funding to support the cloud services.

Question 24 - On a scale from 1 to 4, how important is the Cloud Computing benefit: Performance and flexibility

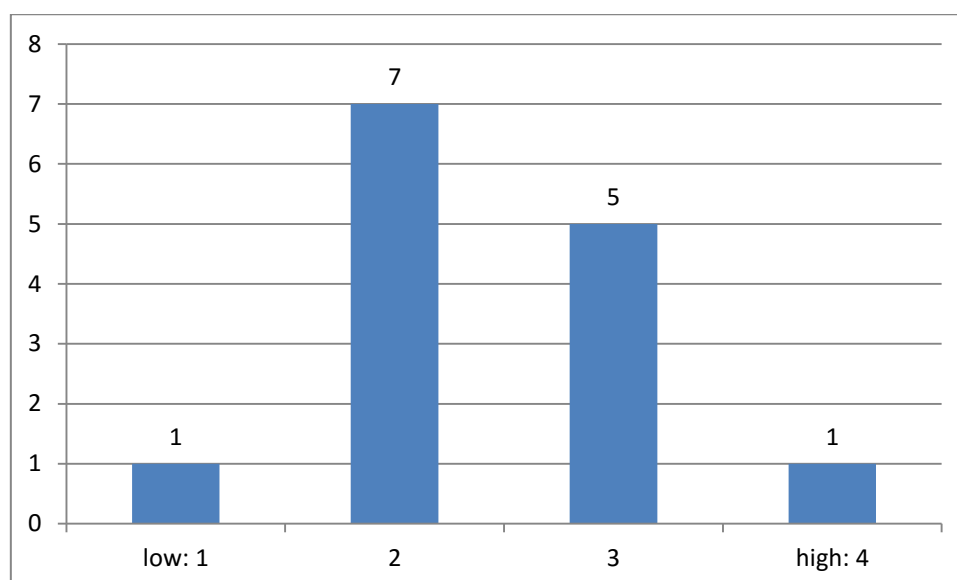


Figure 18 - Question 24 - On a scale from 1 to 4, how important is the Cloud computing benefit: Performance and flexibility

Performance and scalability are important to the respondents, but not highly important.

Question 25 - Would you like to improve a current service of your institution/company by moving to a Cloud-based service?

Below are all the replies. Underlined are the notable replies, most pertinent to the question asked.

- The centre provides services for -omics tools in genetics. The cluster is limited. Also, some parallelization for the statistical package is being implemented, and big matrices, and/or many more users may not be affordable beyond a low limit.
- Our grid computing facilities are closed off to the Internet. We would benefit from cloud-based computing services that are able to integrate cloud-based data streams into our models when running in the cloud.
- Yes, we would like to put all our data and audit services on-line to allow any fishery easier access to the data and tools required to perform MSC audits. Additionally, we would like interoperability around traceability solutions for fish products such that data standards existed to support the secure exchange of data between commercial enterprises, government and INGO's.
- Maritime security applications
- Processing is key for us even if researchers are keen to use supercomputers, cloud solutions need to be investigated.
- We are currently investigating options to move our mapping services into the cloud to improve reliability and access.
- The provision of data to the tuna Regional Fisheries Management Organisations (RFMOs) by each country is currently based on ad-hoc formats defined by each RFMO, which are then converted "internally" into each RFMO database. Each format differs between RFMO and also varies in time, which makes the work difficult. The implementation of cloud-based services to facilitate the provision and processing of the data with transparent tools and equations would be very helpful in that matter.
- RvLAB⁷⁸ (Virtual research environment dedicated to the R statistics package). Ecological Modelling vlab. VRE mainly dedicated to niche modelling.

Question 26 - Do you have difficulty accessing ICT resources? If yes, which ones?

Below are the replies pertinent to the analysis:

- Accessing data in China due to state firewall.
- Yes, experienced people in cloud services and developers with expertise in SaaS applications.

6.3 Interviews

Four interviews were performed with three different types of entities from four different domains; two of the four represent private profit entities.

⁷⁸ <https://portal.lifewatchgreece.eu/>

Below are the interviewees:

1. Mark Luckins, MSC - Fisheries catches traceability/certification/Quality control, Private non-profit sector, IT Director
2. Jean Yves Lebras, CLS - Maritime surveillance (MSC)/safety at sea, Private profit sector, Project Manager
3. J.Barde, IRD - Marine fisheries research, Public sector, Research Engineer in charge of information systems in my research unit and related interoperability issues
4. K Seferis, I2S - Marine aquaculture production and monitoring, Private profit sector, CEO

The interviews were structured to get more qualitative information based on the replies they provided to the questionnaire, and to give them an opportunity to give additional insights of the data needs of the community, this through an “open” conversation.

The insights received are directly reported in Chapter 8 - Findings.

6.4 The BlueBRIDGE Case

BlueBRIDGE, a European project funded under the H2020 framework that began in September 2015. It supports the Fishery and Marine Sciences Community and further develops and exploits the iMarine e-Infrastructure data services for an ecosystems approach to fisheries. This project is a successful example of how to continue (from the iMarine initiative⁷⁹) to cater to the Fishery and Marine Sciences communities.

The objective of the project is to support capacity building in interdisciplinary research communities actively involved in increasing scientific knowledge about resource overexploitation, degraded environment and ecosystem with the aim of providing a more solid ground for informed advice to competent authorities and to enlarge the spectrum of growth opportunities as addressed by the Blue Growth Societal Challenge.

The initiative capitalizes on past investments and uses the proven D4Science infrastructure⁸⁰ that counts over 1500 users, integrates more than 50 repositories, executes around 13,000 models and algorithms per month, and provides access to over a billion records in repositories worldwide, with 99.7% service availability.

BlueBRIDGE is developing innovative services in the following areas:

- Blue Assessment - services for stock assessment and for the generation of unique identifiers for global stocks;
- Blue Economy - services supporting the analysis of socio-economic performance in aquaculture;
- Blue Environment - spatial planning services to identify aquaculture and fisheries infrastructures from satellite imagery;

⁷⁹ <http://www.i-marine.eu/Pages/Home.aspx>

⁸⁰ <https://www.d4science.org/>

- Blue Skills - on-line training services and capacity building on existing training modules for fisheries scientists and other practitioners;
- In addition, exploring how to increase profits and minimize environmental impact with BlueBRIDGE aquafarming services for SMEs.

Aquaculture is one of the pillars of the EU's Blue Growth Strategy⁸¹ and its development can contribute to the Europe 2020 Strategy⁸². Although it represents a relatively small part of the EU economy, it has the potential to boost growth and jobs in EU coastal and inland areas.

In recent years, aqua-farming companies have been competing in an extremely low profit margin environment, thus relying on high sales volume to create adequate profits. This landscape leaves little room for inefficient operations. Another major issue is the environmental impact and the environmental sustainability of the production. Aquaculture, in common with many other sectors, uses natural resources and interacts with the environment on issues of environmental protection. Efficient production management and the development of best practices respond to the above needs. It can dramatically help companies, most of which are SMEs, to improve profitability and minimize environmental impacts.

BlueBRIDGE is developing two new services addressing two relevant problems related to this challenge that build one upon the other:

- Performance evaluation, benchmarking and decision making in aquaculture service: providing capacities for companies to evaluate, benchmark and optimize their performance against best practices and the competition, and to extend the capacity of scientific research communities and policy makers to quantify and comprehend aqua-farming industry operation, ensuring sustainability and development of the sector.
- Strategic Investment Analysis and Scientific Planning and Alerting service: supporting investors and scientists in the efficient identification of strategic locations of interest that meet multifactor selection criteria.

The two new services will be put in practice initially in two domains:

1. A group of aquafarming SMEs, that have been preselected and will be contributing to the benchmarking and evaluation of their production
2. A group of individual stakeholders, not funded by the project, for evaluating potential investment scenarios

The project BlueBRIDGE has recently begun collaborating with EGI through the EGI-Engage project, namely in two Tasks:

1. Task JRA2.3⁸³ (e-Infrastructures Integration) fosters the expansion of the capacity and capabilities of EGI by the integration of its technical solutions with those offered by the e-Infrastructure which serves this community (Fishery and Marine Sciences Data Analysis Sector) through the BlueBRIDGE project. The integration with the gCube/D4Science data and

⁸¹ http://ec.europa.eu/maritimeaffairs/policy/blue_growth/

⁸² http://ec.europa.eu/europe2020/index_en.htm

⁸³ <https://wiki.egi.eu/wiki/EGI-Engage:WP4>

a computational e-Infrastructure will see an update of the federation model of the EGI Federated Cloud.

2. Task SA1.3⁸⁴ (Integration, Deployment of Grid and Cloud Platforms) sets out to deploy and maintain a set of fishery and marine sciences VREs to offer innovative working environments with the as-a-Service paradigm of both gCube and EGI. These facilities are to be offered by exploiting EGI resources, namely the hosting of services and data. A number of representative VREs will be pro-actively created to act as a sort of typical environment ready to use for a class of users (e.g. biodiversity students, data managers). In addition VREs will be specifically created to serve the needs of specific use cases identified during the project. This activity will also support the adaptation of existing applications and data to the VREs.

Following the progress of BlueBRIDGE throughout its lifetime and viewing the results will give EGI valuable insights about how to meet the data needs of the Fishery and Marine Sciences communities.

⁸⁴ <https://wiki.egi.eu/wiki/EGI-Engage:WP5>

7 Findings, Conclusions and Next Steps

The overall objective of this report was to produce valuable input to both EGI technical and non-technical teams that could be analysed and translated into concrete requirements. Therefore the objective of the final section is to provide these findings in a clear, re-usable form. As this report is only a first step in understanding how to better target these communities, a series of next steps have also been provided to ensure that the analysis and content herein will be taken forward.

7.1 Summary of Key Findings

The overall findings of this study are positive. About half of the respondents use cloud-based services, and the others are interested in or plan to move to cloud-based provisioning in the next 2-5 years. This surely gives EGI an indication that the community is ready for services and solutions it has to offer, and perhaps would be interested in co-creating.

Importantly, only one of the respondents is familiar with EGI, which demonstrates this is indeed new territory for EGI and every finding is potentially useful to learn about the community. One important finding is the interest in a new service that would combine different types of cloud services, besides storage and computing infrastructure, with the reuse of data from third parties, software and technical support.

Furthermore, another notable Finding, confirmed through the questionnaire, is the need on behalf of the respondents to access external data, not only from other domains in the Fishery and Marine Sector, but also in other Sectors and related domains, such as geospatial and biodiversity. Access to relevant data in these “other” domains was rated the second highest priority for the respondents, for example Data-as-a-Service (DaaS).

The top-down approach coupled with the bottom-up approach for a market report of this nature is an effective way to introduce the Fishery and Marine Sciences Data Analysis Sector, and give valuable insights into a community relatively unknown to EGI.

The following table summarises the key findings in a “fact” style to allow EGI to easily extract the information for analysis.

Findings

Fishery Exploitation is mainly, if not exclusively, private (individual, SME and industries), whereas Monitoring and Management is mainly public (national and regional governmental/public organizations).

| | |
|---|--|
| Domains where large volumes of data is found. | <p>High interest domains for EGI:</p> <ol style="list-style-type: none"> 1. Marine Fisheries Exploitation and Monitoring, especially for industrial fisheries 2. Marine fisheries research to provide stock assessment, where there is few type of data but with large volume 3. Fisheries catches traceability/ certification/Quality control, where there is a large number of different types of data from different stakeholders 4. Marine environmental research, where there is a need for harmonized data by public companies 5. Marine aquaculture research |
| Marine Aquaculture Production and Monitoring Domain | The stakeholders in this domain are the main providers in aquaculture with different challenges in terms of data storage, processing, dissemination and innovation for data monitoring. |
| Impact of Fisheries on Ecosystem Study – Marine Protected Areas (MPA) Efficiency Domain | A large quantity of ad-hoc data is collected. |
| Maritime Surveillance (MCS)/Safety at Sea Domain | A current challenge is processing data processing coming from different sources. |
| Marine Fisheries Policy Making/Management Domain | The EC is in the process of standardising fisheries reporting formats from all stakeholders (Member States, RFMO, etc.) through UN/CEFACT ⁸⁵ standards and this could create business. |
| Marine Fisheries Exploitation and Monitoring Domain | A main challenge, being data collection. There is a need to store large amounts of data, mostly confidential. Processing and computing needs vary from one institution to another, mainly coming from Research Institutes. Europe is providing an answer with the new Integrated Fisheries Data Management Programme (ex-FLUX). |
| Fisheries Catches: Traceability /Certification / Quality Control Domain | This domain requires a need to store various types of data for long-term duration. Also, facilities are needed to update this data/information easily. The certification processing is highly manual. In regards to traceability, several software suites exist. |

⁸⁵ <http://www.unece.org/cefact>

Marine Fisheries Research Domain

This domain presents typical needs of research: capacity to store large amount of raw data for long-term duration, capacity of computing models and capacity to share/exchange data with colleagues. Marine Fisheries Research is a domain with large quantities of produced data in scattered projects that could benefit from cloud storage solutions.

Marine Fisheries Research Domain

Research data type reflects the large variety of fisheries related studies. Large amount of data can be collected in different formats (data can be stored in Excel tables, in databases (PostgreSQL is very popular in research) or as R tables (R being an open source statistical processing tool)) and from various sources. Ad-hoc formats defined at the study level are implemented. Sharing this information would require harmonization and standardization.

A key element in research is the ability to store all data as it could be reused later on to re-validate the study itself and to exploit new methodologies (on stock assessment).

Usually researchers do not share their raw data. This presents a problem, because when they do, work has to be done to document raw data for exchange. Exchange is basically sharing the existing data in the existing format to another researcher. As a methodology, this is not very efficient as there is no common format to exchange data between researchers.

In Marine Research Institutes tools and processing capacities needs vary from one research department to another. Mostly, ad-hoc solutions are developed under regular programmes or projects funds. There is a common use of open-source solutions (R⁸⁶ for statistical processing, postgresQL for DBMS). IT activities outsourcing could be a source of revenue.

The use of external infrastructure for cost efficiency could be promoted. Revenues will be generated by providing these institutions support to store and process their data, possibly to access more advanced piece of software, and long-term maintenance.

On national/regional level technical capacities to exploit existing tools (bricks of software available in an infrastructure like iMarine) is needed to quickly develop data collection and processing tools for national institutions or RFMO. Funds come from donors, a tender is opened, and a consultancy companies applies: whichever one is more cost efficient (quicker development) likely gets the tender.

⁸⁶ <https://www.r-project.org/>

Fishery and Marine Science Data Analysis Sector comments on types of data (formats, standards, metadata):

- Regarding fisheries data, you have as many formats as you have countries/regional institutions/international organisations. DG MARE with FLUX is in the process of rationalisation.
- The FLUX standard from DG MARE (it should be named with the new name: Integrated Fisheries Data Management Program): all standard formats for data exchange are submitted for validation to UN/CEFACT, an institution in charge of managing and maintaining standard (similar to ISO⁸⁷ (International Organization for Standardization)). Note: There is not much more for the rest of world.
- Typical datasets: Catches/landings/fleet capacity/ fishing effort/CPUE⁸⁸ (Catch per unit effort)/Stocks Assessment Status/Monitoring control and Surveillance data
- Metadata standards:
 - The CWP standard at international level (FAO) for species, fishery commodities, gear type, vessel type. There are standard definition for the datasets: catches/landings/fleet capacities/fishing effort and also CPUE
 - There is a standard for stock description (which includes stock assessment) = FIRMS⁸⁹ (Fisheries and Resource Monitoring System) and is also a standard description of fisheries
- Regional standards: Eurostat and JRC for European Community, FishFrame⁹⁰ for North Atlantic fisheries management (from ICES in Copenhagen)

FINDINGS FROM THE QUESTIONNAIRE

Data worked with is increasingly global (50 percent of respondents) and requires the right technology to manage such data.

The data collected in this Sector is increasing rapidly in size (e.g. from annual national statistical datasets to individual vessel reports), detail (from landings to fully monitored fisheries), precision (e.g. from national EEZ to exact vessel locations) and media (from paper reports to video). The technology thus has to adapt to manage this rapid growth, and include web-services for reference data, statistical packages for time-series analysis, video analysis software, etc.

GIS data is managed by just over 70 percent of the respondents. To note, exploitation and Biological data is managed by 64 percent. Exploitation data is at the beginning of the data value chain, therefore other types of data/information depend on it.

The single top priority in data management is access to quality data in the domain, at just over 70 percent. Interestingly, in second rank is Access to data from related domains (geospatial, biodiversity, etc.) at 57 percent. To note, access to scalable processing capacities was ranked last.

All respondents need access to external data and many from other domains.

⁸⁷ <http://www.iso.org/iso/home.html>

⁸⁸ <http://www.fao.org/docrep/x5685e/x5685e04.htm>

⁸⁹ <http://firms.fao.org/firms/en>

⁹⁰ <http://www.ices.dk/marine-data/data-portals/Pages/RDB-FishFrame.aspx>

Challenges encountered in regards to specific regulations, standards, formats, protocols include:

1. Data security and integrity
2. Colleagues entering data have difficulties to comply
3. Standards cannot fit the richness of biodiversity data, and ad hoc solutions have to be integrated on the fly
4. The export of data in standard formats is a full-time job that requires extra human resources and is very specific
5. There is currently few information transferred to the scientists who work on data (i.e. fisheries data) but do not know much about international standards, regulations, etc.
6. There is a strong need to provide assistance to the research institutes so that they improve their understanding of the evolutions in data tools and formats
7. Availability of appropriate metadata and keeping this linked with the data products
Licencing, referencing of data (especially restrictions on data use), access to data in useable formats
8. Data accessibility, especially socio-economic data, is a big problem, standards to find and integrate data must be further implemented

Main challenges regarding data management in 2-5 years: Data exchange, at almost 80 percent, remains by far the challenge respondents foresee. In second rank (57 percent), respondents chose Dissemination and confidentiality. Data quality (50 percent) is particularly important for this community to perform well. Hosting, Storage and Processing don't worry respondents much.

57 percent of respondents use Cloud-based services, of those (57 percent):

- We can observe an almost even distribution, where all services (storage, hosting, processing) are used.
- The majority, at 37 percent use a private cloud. Ranked second is public (commercial organization) cloud at 27 percent.
- 37 percent (those planning to invest in other services) are looking to use processing services. 31 percent are interested in storage.

Those respondents not using cloud-based services, but interesting in investing in the future, are leaning towards hosting and processing almost in equal percentage (37, 38 respectively), taking priority over storage.

Legal and governance (legal restriction to use cloud-based services, lack of clear roles of ownership) is the top challenge/obstacle recognized for cloud-based services at 57 percent. Budget/Cost, at 50 percent, seems to always be a natural concern, as most of the respondents represent public organizations.

Insights about what the respondents feel the challenges/obstacles (in regards to cloud-based services) are and how they can be overcome include:

- Inertia exists due to legal complications/ambiguity, specific regulations in some industries that prevent, inhibit or constrain uptake and lack of understanding at some senior levels. Some companies may have invested extensively in an on premise estate and moving such arrangement into the cloud can be complex and time consuming as many cloud solutions are limited in some way or another. This can make uptake extremely challenging even when the internal will exists. Finally, time is also required before such investments can be perceived as legacy and the business case for change made.
- A challenge is getting the funds required to maintain such cloud systems, which need to be sustainable when most budgets of research institutions now rely on short-term project (i.e. a few years).
- A second issue would be to share the applications and services with other research institutes and Universities who have similar needs to maximise the interest of the approach and share techniques and processes among partners. This requires to well identify the needs of each potential partner and to implement simple technical tools that do not need high qualifications in computer science.
- A small not-for-profit organization may not be able to afford expensive processing facilities. Therefore, affordable and scalable cloud facilities to run large ecosystem models would be a requirement.
- Long-term is a must to be able to overcome challenges and obstacles.
- An idea to overcome some challenges would be: Network of local solution (Medium size cluster) and accessing to several clouds.
- The challenge to use some processing services delivered by the cloud is more in having technical advices and making the use of such service seamless for researchers.

Perception of the Cloud Computing benefit:

- “Facilitate analysis of cross-domain information through harmonization and standardization” is high. This benefit refers to:
 1. Harmonize storage: Enable better systems (bring data to cloud and reduce physical separation making it easier to access)
 2. Harmonize processing: To enable scalable analysis (data repositories can be huge - combining data from two domains requires large and flexible processing capacity)
 3. Harmonize reference data: New data (e.g. observations) can be harmonized to comply with geospatial or other reference datasets (reference data on cloud could be easier to reference - harmonization of content)
- “Cost saving” is not particularly high.
- “Performance and flexibility” is not particularly high.

Insights about how important cost is when choosing Cloud-based services include:

- Public research bodies say reducing costs is unfortunately important. It would be fine to get more services with the same budget. However, cloud solutions for Southern countries are still irrelevant in some cases due to string limitations of bandwidth or stability of Internet access.
- One key issue with high value cloud service offerings is that they are not affordable for SMEs. Once affordability is addressed, cloud economics move organizations from a CAPEX to OPEX cost, which is not always wanted in a not-for-profit/charity space.
- Cloud providers are very distant from customers and define services in terms of "storage" etc. when businesses need a "full service story" (i.e. advice, consultancy, training, and help.)
- The cost of long-term needs will be scrutinized in detail. The costs for one-shot or short-term needs will be less important depending on the importance or urgency of the work.
- A case presented: many of the datasets collated are confidential and many countries would be reluctant to report data if the storage, processing and dissemination were cloud-based rather than data stored in-house. For this reason, some entities have not fully explored or considered options for cloud-based services.
- It is important that the cost is scalable and not prohibitive. There are also challenges in using cloud services on a project basis, where there is a fixed time period of funding to support the cloud services.

Some respondents would like to improve a current service of their institution/company by moving to a Cloud-based service:

- Seeking to increase the limit of Users, as many more users may not be affordable.
- Grid computing facilities are closed off to the Internet. We would benefit from cloud-based computing services that are able to integrate cloud-based data streams into our models when running in the cloud.
- Interoperability around traceability solutions for fish products such that data standards existing to support the secure exchange of data between commercial enterprises, government and INGOs.
- Looking to improve reliability and access.
- The provision of data to the tuna Regional Fisheries Management Organisations (RFMOs) by each country is currently based on ad-hoc formats defined by each RFMO, which are then converted "internally" into each RFMO database. Each format differs between RFMO and also varies in time, making the work difficult. The implementation of cloud-based services to facilitate the provision and processing of the data with transparent tools and equations.

7.2 Conclusions and Next Steps

As mentioned, the overall objective of this report was to produce valuable input to both EGI technical and non-technical teams that could be analysed and translated into both concrete technical and business requirements.

Therefore, the following month of activities will focus on bringing together the community experts (authors of this report) with the EGI User Community Support, Operations and Strategy Teams in order to translate each finding into the appropriate channel, whether technical requirements into the iMarine integration work part of EGI-Engage WP4 (Platform Integration - gCube) and WP5 (Operating VREs for the iMarine user community) and/or business requirements for the integration or creation of new services i.e. introduction of a data-as-a-service (DaaS) within EGI.

Formally, each requirement/recommendation will be added as a dedicated ticket with the EGI Tracker tool (RT)⁹¹ to be handled by the most appropriate team or individual. Initial work has already begun to understand the structure of the RT template and the information required, which led to the identification of a few modifications that will need to be agreed and addressed.

The results of this work will be reported in the first Periodic Report and accounted for in a dedicate EGI-Engage Activity metric - M.NA2.Industry.3 “Number of requirements gathered from market analysis activities”.

It will be essential that all analysis consider these findings by exploring how EGI solutions and services can meet the data needs expressed, now and in the future. This means focusing on the data needs and challenges of the stakeholders as a starting point, getting to know this community through forming relationships and developing a data-as-a-service (DaaS) proposition in collaboration with other initiatives and projects (e.g. BlueBRIDGE).

⁹¹ <http://rt.egi.eu>

8 References

Ocean Research in Horizon 2020: The Blue Growth potential,

[http://www.europarl.europa.eu/RegData/etudes/STUD/2015/518775/IPOL_STU\(2015\)518775_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2015/518775/IPOL_STU(2015)518775_EN.pdf)

Towards European Ocean Observation,

https://ec.europa.eu/research/infrastructures/pdf/toward-european-intagrated-ocean-observation-b5_allbrochure_web.pdf

The State of World Fisheries and Aquaculture, 2014, <http://www.fao.org/3/a-i3720e/index.html>

Facts and figures on the Common Fisheries Policy,

http://ec.europa.eu/fisheries/documentation/publications/pcp_en.pdf

EGI 2020 Strategy, <https://documents.egi.eu/document/2417>

An Open Science Cloud to Realize the Data Commons,

<https://documents.egi.eu/public/RetrieveFile?docid=2575&version=5&filename=OpenScienceCloud-EGI-v1.pdf>

Appendix A

As specified in Chapter 4, focus was given to certain Domains presenting opportunities to EGI. Here below is the description of the domains excluded from the study, but which contribute to an overview of the Fishery and Marine Sciences Data Analysis Sector.

Illegal, Unreported and Unregulated Fishing (IUU)

Illegal, Unreported and Unregulated Fishing (IUU) is a worldwide concern and a priority. Work has been done to tackle this concern:

1. The FAO provided recommendations on IUU to Member States in the Fisheries Code of Conduct in 1995.
2. An International Plan of Action to Prevent, Deter and Eliminate IUU fishing (IPOA-IUU) was developed in 2001 by FAO and its Member States.
3. The EC addressed this issue in the 2008 and 2009 regulations

IUU has a direct impact on fisheries sustainability. It is not limited to high seas but is also a concern for coastal states. Consequences are economical in developed countries but can be on population subsistence in developing countries (food security issues).

Fighting IUU is a complicated task as it requires the adoption of a vessel monitoring tool (VMS), which is costly and not adapted to artisanal fisheries, national MSC resources and regional MCS coordination. The International MCS network was created in 2001 to facilitate such coordination. But this network is on a voluntary basis and operates informally.

A regional MCS network for West Africa was supported by an EU funded project in 2010-2013, but faced difficulties to coordinate MCS activities from the different countries (problem of local resources to feed to regional network).

Estimating IUU is not an easy task, as it requires cross-domains information and data, including:

- MSC national data which is not always publicly available
- Comparable reported landed and sales data
- IUU sighting activities and standard reporting that does not exist yet

Although being a worldwide concern, IUU fishing assessment relies on very heterogeneous data coming from a lot of different scattered sources.

Coastal - Maritime Tourism

This domain is for some countries of crucial economic importance. Tourism includes related infrastructure (hotels, marinas, restaurants, beaches) that can have an impact on marine environment and activities such as sport fishing, scuba diving, sailing, speedboats, with impacts such as pollution, pressure on local fish resources, etc. Few data and studies are made on these

impacts. Globally, tourism is managed by independent private businesses, which do not generate a large amount of data.

Tourism activities mainly produce exploitation data (income, number of staff, production, and turn over) for private/internal use.

Oil / Gas Exploration and Exploitation and Infrastructure (Platforms, Pipes, Storage)

This domain is also for some countries of crucial economic importance. Mainly exploitation is done in high sea, although off shore could be found close to the shore in some African countries (Nigeria for instance). This domain encompasses all activities related to sea fossil fuel exploitation that is mainly static, produces limited data (exploitation statistics), and consumes limited data (weather forecasts, staff rotation schema).

Maritime transport does not impact the ocean (except in rare occasions with an accident resulting in catastrophic pollution). It does not report except on its position and does not interact with maritime resources. It produces one type of data (position), and consumes limited data (weather forecasts).

Piracy

Marine piracy is a worldwide issue with a serious impact on maritime commercial routes and marine fisheries. Although of crucial importance for safety at sea with an impact on fisheries (dramatic decrease of fisheries in certain parts of Indian Ocean for instance), such impact is not yet monitored. NATO is working on Piracy and express's interest and need for secured cloud services.

Maritime Transportation and Infrastructure (Main Sea Routes, Ports and Lighthouses)

This domain encompasses all commercial and non-commercial marine transportation activities and related infrastructure. Data produced are maritime traffic: number of vessels at sea, quantities/values of good transported and landed in port, and number of accidents.

Offshore wind turbines and related sustainable energy infrastructure (cables)

This relatively new domain includes all offshore wind turbines and new undersea or tidal turbines. Data produced includes: number of turbines and potential power production available/actual production of these turbines per day (to be compared with winds to measure turbines efficiency), position of main marine power cables from turbines to land, and number of kilometres of these cables (gives an indication on loss of energy).