



EGI-Engage

Report data sharing policies and legal framework in fishery and marine sciences

D2.6

| | |
|------------------------|---|
| Date | 30 March 2016 |
| Activity | NA2.3 |
| Lead Partner | FAO |
| Document Status | FINAL |
| Document Link | https://documents.egi.eu/document/2699 |

This document describes constraints and requirements given by Intellectual Property Rights (IPRs) and overall legal frameworks that may hinder the sharing of data among different institutions in the fishery and marine sciences communities; instructions to data providers and consumers on how to describe and access the life-cycle of data and contents; recommendations of how to support legal interoperability and confidentiality of data owners.



This material by Parties of the EGI-Engage Consortium is licensed under a [Creative Commons Attribution 4.0 International License](#).

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 CC 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: | Eise van Maanen Anton Ellenbroek Vassilis Protonotarios (Agroknow) | FAO/NA2.3 FAO/NA2.3 | 29/02/2016 |
| Moderated by: | Malgorzata Krakowian | EGI.eu/NA1 | |
| Reviewed by | Mariusz Sterzel Matthew Viljoen Jens Jensen | CYFRONET EGI.eu STFC | 14/03/2016 |
| Approved by: | AMB and PMB | | 30/03/2016 |

DOCUMENT LOG

| <i>Issue</i> | <i>Date</i> | <i>Comment</i> | <i>Author/Partner</i> |
|--------------|-------------|--|--|
| v.0 | 25/11/2015 | Draft ToC | Anton Ellenbroek (FAO) |
| v.1 | 04/02/2016 | Interviewed Julien Barde (IRD), Nadia Nardi (ENG), Aureliano Gentile (FAO), Gianpaolo Coro (CNR); conducted literature review and start of desktop study; connected with relevant interviewees; connected with senior legal experts. | Eise van Maanen (FAO), Anton Ellenbroek (FAO) |
| v.2 | 09/02/2016 | Interviewed Jeroen Steenbeek (EcoPath), Johannes Keizer (FAO) and continued writing ch 4 | Eise van Maanen (FAO), Anton Ellenbroek (FAO) |
| v.3 | 12/02/2016 | Processed interviews and continued work on ch 4 | Eise van Maanen (FAO), Anton Ellenbroek (FAO) |
| v.4 | 17/02/2016 | Continued overall writing | Eise van Maanen (FAO), Anton Ellenbroek (FAO) |
| v.5 | 19/02/2016 | Added the interviews | Eise van Maanen (FAO), Anton Ellenbroek (FAO) |
| v.6 | 24/02/2016 | Wrote introduction and rewrote chapters 3,4 and 6 | Eise van Maanen (FAO), Anton Ellenbroek (FAO) |
| v.7 | 26/02/2016 | Wrote exec summary and finalised first draft version | Eise van Maanen (FAO), Anton Ellenbroek (FAO) |
| v.8 | 29/02/2016 | Draft version | Eise van Maanen (FAO), Anton Ellenbroek (FAO) |
| v.9 | 02/03/2016 | Wrote and update the information in the Chapter 5 about the data sources analysis | Vassilis Protonotarios |
| v.10 | 22/03/2016 | Updates on Chapter 9 and integration of reviewers' recommendations | Vassilis Protonotarios |
| v.11 | 25/03/2016 | Final version | Eise van Maanen (FAO), Anton Ellenbroek (FAO), Vassilis Protonotarios (Agroknow) |
| FINAL | 29/03/2016 | Quality review for final approval | Sy Holsinger (EGI.eu) |

TERMINOLOGY

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

Contents

| | |
|---|----|
| Table of Figures | 6 |
| Executive summary..... | 7 |
| 1 Introduction..... | 9 |
| 2 Background..... | 11 |
| 2.1 Objectives | 11 |
| 2.2 Scope | 11 |
| 2.3 Intended Audience | 12 |
| 2.4 Approach..... | 12 |
| 2.5 Data sharing in the fishery and marine sciences sector..... | 12 |
| 3 Landscape / Seascape | 14 |
| 3.1 Legal interoperability instruments | 14 |
| 3.1.1 Legal interoperability..... | 14 |
| 3.1.2 Legal interoperability instruments..... | 18 |
| 3.2 Data structure and size | 23 |
| 3.3 Global initiatives | 24 |
| 3.4 EU initiatives | 28 |
| 4 Consultation on Stakeholders and Data Sources | 31 |
| 4.1 Identification of Stakeholders | 31 |
| 4.2 Stakeholder classification..... | 32 |
| 4.3 Indicative key data stakeholders | 33 |
| 4.3.1 IODE – International Oceanographic Data and Information Exchange..... | 33 |
| 4.3.2 National Oceanic and Atmospheric Administration (NOAA) Centers | 35 |
| 4.3.3 World Meteorological Organization (WMO) | 37 |
| 4.3.4 SeaDataNet..... | 37 |
| 4.3.5 International Council for the Exploration of the Sea (ICES)..... | 37 |
| 4.3.6 Australian Ocean Data Network..... | 38 |
| 4.3.7 iMarine | 38 |
| 4.3.8 Eurostat | 39 |
| 4.4 Fisheries and Marine Data Coverage..... | 39 |
| 4.5 Analysis of Indicative Fisheries and Marine Data Sources | 41 |
| 4.5.1 IODE Ocean Data Portal..... | 41 |

| | | |
|--------|--|----|
| 4.5.2 | World Ocean Database | 41 |
| 4.5.3 | OBIS (Ocean Biogeographic Information System) | 42 |
| 4.5.4 | Marine Geoscience Data System (MGDS) | 42 |
| 4.5.5 | Marine Environmental Data & Information Network (MEDIN) | 43 |
| 4.5.6 | Australian Institute of Marine Science (AIMS) | 43 |
| 4.5.7 | Integrated Marine Observing System (IMOS) | 44 |
| 4.5.8 | FAO List of Species for Fishery Statistics Purpose (ASFIS) | 44 |
| 4.5.9 | FAO Fisheries and Aquaculture Department..... | 44 |
| 4.5.10 | General Fisheries Commission for the Mediterranean (GFCM)..... | 45 |
| 4.5.11 | Joint Research Centre (JRC) | 45 |
| 4.5.12 | Global Biodiversity Information Facility (GBIF) | 46 |
| 4.6 | Content Analysis of the Fisheries and Marine Data Sources..... | 46 |
| 4.6.1 | National Climatic Data Center..... | 46 |
| 4.6.2 | SeaDataNet..... | 48 |
| 4.6.3 | OBIS | 48 |
| 4.6.4 | MGDS..... | 48 |
| 4.6.5 | MEDIN | 49 |
| 4.6.6 | ICES..... | 50 |
| 4.6.7 | AIMS Data | 50 |
| 4.6.8 | AODN Ocean Data Portal | 51 |
| 4.6.9 | IMOS..... | 51 |
| 4.6.10 | FAO ASFIS | 51 |
| 4.6.11 | FAO Fisheries and Aquaculture Department..... | 52 |
| 4.6.12 | FAO GFCM | 52 |
| 4.6.13 | Joint Research Centre (JRC) fisheries data collection | 52 |
| 4.6.14 | GBIF | 53 |
| 4.7 | Data flows | 53 |
| 4.7.1 | Data usage and citation | 53 |
| 4.7.2 | Data management | 54 |
| 4.7.3 | Infrastructure support to data flows..... | 54 |
| 4.8 | Legal Analysis of the Fisheries and Marine Data sources | 55 |
| 4.8.1 | IODE..... | 55 |
| 4.8.2 | NOAA / NCEI Centers | 56 |

| | | |
|--------|---|----|
| 4.8.3 | SeaDataNet..... | 56 |
| 4.8.4 | OBIS..... | 57 |
| 4.8.5 | MGDS..... | 57 |
| 4.8.6 | MEDIN | 58 |
| 4.8.7 | ICES..... | 58 |
| 4.8.8 | AIMS | 59 |
| 4.8.9 | IMOS & AODN..... | 59 |
| 4.8.10 | FAO ASFIS | 60 |
| 4.8.11 | FAO Fisheries and Aquaculture Department..... | 60 |
| 4.8.12 | FAO GFCM | 60 |
| 4.8.13 | JRC..... | 61 |
| 4.8.14 | iMarine | 61 |
| 4.8.15 | Eurostat | 61 |
| 4.8.16 | GBIF | 62 |
| 4.9 | The bigger picture | 62 |
| 4.10 | Stakeholder Findings Representing Opportunities for EGI | 63 |
| 5 | Interview Analysis..... | 66 |
| 6 | Use Case: BlueBRIDGE VRE in support of a Stock assessment Regional Database | 71 |
| 6.1 | WECAFC Stock Assessment VRE | 74 |
| 6.2 | The Global Record of Stocks and Fisheries VRE and WECAFC data | 76 |
| 6.2.2 | VRE specific data policy and legal challenges, functionalities | 77 |
| 6.2.3 | Protected Area Impact Maps VRE and WECAFC spatial data | 78 |
| 6.2.4 | VRE specific data policy and legal challenges, functionalities related to WECAFC potential products | 78 |
| 7 | Findings and Opportunities | 79 |
| 8 | Conclusions..... | 84 |
| 9 | Bibliography..... | 87 |

Table of Figures

| | |
|--|----|
| Figure 1: Indicative view of the marine and fisheries key stakeholders | 32 |
| Figure 2: Indicative view of the marine and fisheries data stakeholders' ecosystem | 33 |
| Figure 3: Mapping of the IODE network..... | 34 |
| Figure 4: NOAA Centers and data services of interest to this report | 35 |
| Figure 5: Structure of the Australian Ocean Data Network..... | 38 |
| Figure 6: Flow of data and information between various actors..... | 40 |
| Figure 7: Indicative view of the ecosystem of data stakeholders in marine and fisheries..... | 41 |
| Figure 8: The AIMS data portal | 50 |

Table of Tables

| | |
|---|----|
| Table 1: CC licenses (<i>adapted from Korn & Oppenheim, 2011</i>) | 22 |
| Table 2: MGDS data sources and services..... | 42 |
| Table 3: NCDC datasets..... | 47 |
| Table 4: Controlled vocabularies used by the MEDINE | 49 |
| Table 5: List of people interviewed in the context of this analysis..... | 66 |

Executive summary

The EGI-Engage project dedicated this document to produce an introduction into the legal aspects of data sharing with a focus on the fishery and marine sciences sector. The main objective of this report is to stimulate legal interoperability of data sharing within the fishery and marine sciences sector. In the information society, legal interoperability drives competition, innovation, economic growth and trade (Palfrey and Gasser, 2012). Legal interoperability thus intends to provide clarity for the use, reuse and sharing of data.

This document has been prepared in conjunction with D2.7 “Market Report on the Fishery and Marine Sciences Data Analysis Sector”. These deliverables pave the road for EGI-Engage to enter the fishery and marine sciences sector. However, this document focuses more on providing an analysis of the legal barriers of data sharing within the sector and legal interoperability instruments that may be used to take away these barriers.

The intended audience consists of the EGI community and experts working with research data in the fishery and marine sciences sector. This report mostly targets people directly involved in data collection, management and analysis, but it also serves as an introduction for practicing lawyers in data sharing.

This report assumes legal interoperability to be: *the compatibility of legal rights, terms, and conditions of databases from two or more sources so that the data may be combined and integrated by any user, without further permission and without compromising the legal rights of any of the data sources used.*

The approach taken in this study is both top-down and bottom-up. The top-down approach consists of a desk study analysing the current legal interoperability instruments of data sharing and assessing relevant initiatives working on data sharing. The main international legal methods to foster data sharing are *standardisation, harmonisation and mutual recognition*. The selection of legal interoperability instruments discussed in this study are *waivers, common-use licenses, contracts and terms of use*.

The bottom-up analysis is conducted through interviews with people with data sharing responsibilities from FAO, PLAZI, iMarine, IRD, ICES, Ecopath and others. The objective of the interviews was to identify legal barriers to share data in fishery and marine sciences across institutes and communities. The interviews exemplified the following legal barriers: opaque legal frameworks to share data; lack of resources; lack of incentives to publish, and tradition to withhold data; citation should be standardised through a European citation guide; public institutions that funded data collection should keep copyright; data sharing is all about trust; and more attention should be given to normative standards. There was a wide consensus among the interviewees that infrastructure services that support a legal framework for data sharing is much needed. EGI services can have an important role in implementing and supporting such a framework.

The BlueBRIDGE project, funded under the Horizon 2020 framework programme, was selected as a best-practice case and presented in this deliverable as it aims to enable different stakeholders to

provide informed advice to competent authorities through a marine data e-infrastructure (based on the outcomes of the iMarine initiative). In this context, a use case studied by BlueBRIDGE is presented as it also concerns legal interoperability of data aggregated from various sources.

1 Introduction

The United Nations Universal Declaration of Human Rights posited already in 1948 the access to information as a fundamental human right¹. The volume and complexity of transborder data flows increased massively since 1948, and particularly in recent years. Data protection due to the increase of the value of data (as stated in the Deliverable D2.7 “Market report on the Fishery and Marine Sciences Data Analysis Sector”²) has spurred a legal movement to develop a legal culture of data management with its own set of rules. This evolved into the current situation where data protection is subject to a maze of rules, in a variety of jurisdictions.

With the growing volume of transborder data flows, global privacy rules should be available. Unfortunately, data protection laws are very different in various regions of the world. A harmonisation of these rules is not expected in the near future, but the costs of non-interoperable laws in a highly networked world are high and will increase even more.

EGI-Engage aims to accelerate the implementation of the Open Science Commons by expanding the capabilities of a European backbone of federated services for compute, storage, data, communication, knowledge and expertise, complementing community-specific capabilities³.

This document works towards this aim through WP2 Strategy, Policy and Communications under the task NA2.3. The objective is to analyse data sharing policies and legal aspects of data sharing, with a focus on the fishery and marine sciences sector. It seeks to:

1. Explore legal barriers in sharing fishery and marine sciences datasets;
2. Deliver a framework of legally relevant instructions for data sharing to data providers and consumers;
3. Validate a use case with a selected community of a regional database targeting fisheries’ productivity;
4. Devise a context where infrastructure support to processing a mix of public and non-public datasets results in improved data availability;
5. Advice on how the legal interoperability is best supported through infrastructure security.

The main objective of this report is to stimulate legal interoperability of data sharing within the fishery and marine sciences sector. In the information society, legal interoperability drives competition, innovation, economic growth and trade (Palfrey and Gasser 2012, 182). Legal interoperability thus intends to provide clarity for the use, reuse and sharing of data. Transparency, establishment and promotion of legal interoperability will perform a highly valuable service towards the aim of open data.

This document is important to EGI-Engage and introduces EGI to the fishery and marine sciences sector. There are many reasons why EGI should be interested in this community. First of all, there are opportunities to be taken since the EU has an important presence in the fishery and marine sciences sector generating an income of €7.2 billion (Nardi, 2016). Second, EGI can foster the

¹ <http://www.un.org/en/universal-declaration-human-rights/>

² <https://documents.egi.eu/document/2700>

³ <https://www.egi.eu/about/egi-engage/>

sustainability of the fishery and marine sciences sector through data sharing. This will enhance data analysis and research on optimal fishing capacity from both an economic and environmental perspective.

This document begins with a description of the background, where information covering the objectives of this document, scope, intended audience and overall approach are discussed.

The second chapter is on the landscape/seascape of this sector. This chapter includes a list of relevant legal interoperability instruments, data structure and size, global and EU initiatives operative in the legal interoperability of data sharing and EGI and partners' existing policies.

The stakeholder analysis in the next chapter identifies and characterises the stakeholders. The focus of this stakeholder analysis is the fishery sector.

Chapter six includes a best-practice case, i.e. the BlueBRIDGE case. This case will be discussed in depth to learn the best-practice aspects, but also the areas of improvement that BlueBRIDGE faces.

Finally, this study ends with a chapter on the findings and opportunities for EGI to improve legal interoperability of data sharing in the fishery and marine sciences sector.

2 Background

This chapter provides the background of this study stating the objectives, scope, intended audience and approach.

2.1 Objectives

This study is part of the second work package of EGI-Engage (EGI-Engage: WP2 (NA2) Strategy, Policy and Communications) with its main purpose to steer the consolidation and growth of the EGI community by developing a strategy towards the Open Science Commons vision and ensure the engagement of all stakeholders. This study follows up on the Deliverable D2.7 *“Market report on the Fishery and Marine Sciences Data Analysis Sector”*.

This study will analyse data sharing policies and legal aspects with a focus on the fishery and marine sciences. The objectives are to:

- 1) Explore legal barriers in sharing fishery and marine sciences datasets (raw and processed data) with other institutions and communities, which would prevent research and business opportunities;
- 2) Deliver a framework of legally relevant instructions to data providers and consumers on how to describe their data, the access to this data, and the lifecycle of data and contents and / or of parts thereof in an infrastructure;
- 3) Validate a use case with a selected community of a regional database targeting fisheries' productivity;
- 4) Devise a context where infrastructure support to processing a mix of public and non-public datasets results in improved data availability whilst respecting legal dissemination boundaries;
- 5) Advise on how the legal interoperability is best supported through infrastructure security, especially where storage and access arrangements are required (e.g. to support confidentiality needs of data owners). This issue will be done by working with Regional Fishery Bodies on the development of regional databases in support to stock assessment and fishery management, or discussing value of public data in data value added chain with its community partners.

In this way, this study seeks to promote a culture of openness and sharing of research data among research communities.

2.2 Scope

This study will give an introduction in data sharing policies and legal frameworks within the fishery and marine sciences sector, which includes fisheries⁴, aquaculture⁵ and maritime sectors.

⁴ A unit determined by an authority or other entity that is engaged in raising and/or harvesting fish. Typically, the unit is defined in terms of some or all of the following: people involved species or type of fish, area of water or seabed, method of fishing, class of boats and purpose of the activities.

The scope of this document is limited to rules pertaining to the access and reuse of data, with a focus on the ability to integrate those data in research purposes. The focus is on facilitating the sharing, access and reuse of data that are either produced primarily in, or with funding by the public sector.

Legal interoperability will be the main focus of this report. In this context, the focus will be on legal interoperability of research data in the fishery and marine science sector.

Although the definition of legal interoperability has been subject to much debate; this study defines legal interoperability as *the compatibility of legal rights, terms, and conditions of databases from two or more sources so that the data may be combined and integrated by any user, without further permission and without compromising the legal rights of any of the data sources used*.

This study provides an insight into existing public law statutory, regulatory, and policy approaches, as well as private law instruments that may be used to share data. However, the findings of this study are restricted to a broad outline of data sharing policies and legal frameworks within the fishery and marine science sector, and will not mention codified law unless it is absolutely required.

2.3 Intended Audience

The intended audience consists of the EGI community and all experts working with research data in the fishery and marine sciences sector. This report mostly targets people directly involved in data collection, management and analysis, but it also serves as a useful introduction for practicing lawyers in data sharing. Institutes with the facilitation of (legal) interoperability of databases as core activity such as BlueBRIDGE and RDA are also intended audience.

2.4 Approach

The approach chosen reflects the approach of EGI's Market Report on the Fishery and Marine Sciences Data Analysis Sector (Nardi, 2016). At first, this report will analyse the landscape through a top-down desk study. This desk study will describe existing legal interoperability instruments, and initiatives engaged in data sharing policies and legal frameworks. Key materials include CIGI's Legal Interoperability as a Tool for Combatting Fragmentation (2014) and RDA's Implementation Guidelines for the Principles on the Legal Interoperability of Research Data (2015). Secondly, this report will take a bottom-up approach by analysing first-hand experience from interviews with stakeholders to gain insights on data sharing in reality. The BlueBRIDGE project will serve as a best practice example, which serves the fishery and marine sciences data analysis domains.

2.5 Data sharing in the fishery and marine sciences sector

Data sharing in the fishery and marine sciences sector is subject to many challenges; this subchapter provides a brief background on these challenges.

⁵ The farming of aquatic organisms including fish, molluscs, crustaceans and aquatic plants. Farming implies some form of intervention in the rearing process to enhance production, such as regular stocking, feeding, protection from predators, etc. Farming also implies individual or corporate ownership of the stock being cultivated.

Current legal frameworks such as intellectual property rights (IPR) and data protection are developed in another technical environment, and are mainly directed to a print environment. The IPR and data sharing frameworks have been updated but clearly struggle with the digital environment. An example is the Data Protection Directive,⁶ which has been developed in 1995, i.e. the early computer age when databases were easily traceable, small and clearly separated. This centrepiece of existing EU legislation on personal data protection is currently being updated through the General Data Protection Regulation (GDPR). The GDPR states that ‘rapid technological developments have brought new challenges for the protection of personal data. The scale of data sharing and collecting has increased dramatically’ (European Commission, 2012). Updating EU legislation is a tedious and lengthy process, and will therefore tend to innovate at a slower pace than technology.

The IPR framework grants the creator of original work exclusive rights for its use and distribution. The objective is to stimulate investment and innovation; but it can be argued that the IPR framework nowadays rather hampers the development of the data economy. The data protection framework aims to protect privacy of personal data and is thus an important component of EU privacy and human rights law. Technological innovations such as ‘big data’⁷ challenge the data protection framework.

The size of datasets, and use of data from many sources complicates the traceability, transparency and access rights to data. This exacerbates when these data sources originate from multiple jurisdictions. IPRs and data protection can thus cause isolated data sources and increase the transaction costs of gathering data due to obtaining all the necessary licenses for each source.

Big data processing also poses major problems for the data protection framework. Basic assumptions within this framework such as distinction between personal data and anonymous data become difficult with big data collection and processing. The rights concerning access and transparency also become more complex with big data (Lammerant et al., 2014).

Sharing of data on a global level is particularly in the fishery and marine sciences sector of the utmost importance, to prevent vulnerable and endangered species from extinction. Incentives to withhold data are legal, economic and strategic. The legal incentives to withhold data will be discussed in chapter four of this study. The economic incentives originate at fishing vessels that do not want to reveal their spatial and temporal fish catch data; this after all reveals their business strategy. Another economic incentive to withhold data exceeds the national level. Fishery of some species have become an industry of significant size, an example is the Bluefin Tuna which provides an annual revenue of more than \$500 million for seafood companies. Individuals involved in data sharing on a national level are thus incentivised to protect these industries from quota, by withholding fish catch data.

One of the main problems of data sharing in this sector is legal interoperability. This term will be explained in the next chapter.

⁶ Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data

⁷ Computational process of discovering patterns in large data sets.

3 Landscape / Seascape

There are many benefits to be derived from open data. Fundamentally, sharing public data without any restrictions has proven to be economically beneficial, creating value multiple times and providing much greater returns on the public investment than restrictive approaches (GEO, 2015). Furthermore, open data is social welfare maximising, and has several side-effects such as diverse reputational benefits for data owners. Placing data in the public domain also facilitates the education of future generations significantly and will improve decision-making and transparency in government and society.

This chapter describes the land- and seascape of data sharing with a focus on the fishery and marine sciences sector. The first subchapter briefly describes all relevant legal interoperability instruments and the data structure and size within the fishery and marine sciences domain. To follow is an introduction to the global, EU and EGI initiatives engaged in data sharing policies and legal frameworks.

3.1 Legal interoperability instruments

3.1.1 Legal interoperability

The term interoperability is increasingly being used within information and data management, its meaning is however somewhat ambiguous. According to the online Cambridge Dictionary interoperability is defined as *‘the degree to which two products, programs, etc. can be used together, or the quality of being able to be used together’*⁸. However, this report adopts the definition of interoperability of Paul Miller (2000):

to be **interoperable**,

one should actively be engaged in the ongoing process of ensuring that the systems, procedures and culture of an organisation are managed in such a way as to maximise opportunities for exchange and re-use of information, whether internally or externally.

There are many aspects of interoperability such as technical, semantic and political interoperability. Although technical interoperability is still a significant challenge in the fishery and marine science data analysis sector, this report focuses on legal interoperability.

The relations between interoperability and the law are many, complex, and entangled. The law can help establish, adjust, or maintain interoperability (Palfrey & Gasser, 2015). At the same time, interoperability is also a feature of the legal system itself, i.e. legal interoperability.

As mentioned in Chapter 3, this report assumes **legal interoperability** to be:

the compatibility of legal rights, terms, and conditions of databases from two or more sources so that the data may be combined and integrated by any user, without further permission and without compromising the legal rights of any of the data sources used.

⁸ <http://dictionary.cambridge.org/dictionary/english/interoperability>

Legal interoperability may occur either within the legal system of a single nation-state or across nations (international legal interoperability), which magnifies the complexity even more. Legal interoperability is mainly an issue of cross-border coherence of normative order, but procedural aspects can also play a role (Weber, 2014).

Legal interoperability becomes even more relevant when merging data from different databases. After all, when data are combined from multiple sources the resulting dataset will incorporate the accumulated restrictions imposed by all sources (RDA-CODATA, 2015). The fewest restrictions in parent datasets, therefore result in the fewest restrictions in derivative datasets.

There are several options for increasing levels of legal interoperability. The point is not to make all legal systems the same, but rather to make them compatible. Legal interoperability sometimes means that new laws should be implemented or existing laws adjusted or reinterpreted to achieve this interoperability (Palfrey and Gasser 2012, 178-179). It is thus not necessary for countries to distribute all legislative authority to e.g. the United Nations and to be deprived of sovereignty.

In order to enable the maximum degree of legal interoperability, access to and reuse of research data⁹ should be either unrestricted by default or otherwise be granted to users with the fewest limitations possible. But the maximum degree is not inherently optimal.

Institutions need to aim for an optimal level of interoperability among legal systems. Too high a level of legal interoperability would cause difficulties in the management of the harmonised rules and disregard of social and cultural differences, while too low a level could present challenges for efficient (social or economic) interaction. A fragile equilibrium must be balanced out between legal fragmentation (sovereign states should be able to develop national law systems) and legal interoperability (Gasser, 2015).

Data in the public domain ensures there are no restrictions, and would thus be the preferred option from an open data perspective. However, this freedom of access and reuse should be balanced against legitimate interests to withhold or restrict such data, or control the use of data. These legitimate interests vary per jurisdiction and situation but this study assumes the following categories:

- 1) Intellectual property rights
 - a. Copyright
 - b. Database protection right
 - c. Restrictive contracts and licenses
- 2) National security laws
- 3) Confidentiality laws and policies
- 4) Protection of endangered species
- 5) Individual contracts or use agreements

These legitimate interests guide the formulation of the legal interoperability instruments.

As mentioned before, equilibrium must be balanced out between the public interest on full legal interoperability and thus free information exchange, against the conflicting legitimate interests.

⁹ Research data are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings

This subchapter will define these countervailing interests and provide some guidance to navigate these restrictions, and subsequently discuss the legal interoperability instruments.

Copyright

Copyright is a legal right that grants the creator of original work exclusive rights for its use and distribution for a limited amount of time. Copyright can exist on individual data as well as over a dataset or database as a whole.

Whether copyright protection exists for a certain dataset, is not easy to determine. The criterion of originality infers that copyright does not apply to facts, ideas or concepts. However, research data mostly refer to facts and are presented in a standardised form that shows no originality or creativity and thus cannot qualify for a copyrighted work (Patterson et al., 2014). In many cases a claim for copyright for research data thus lacks any legal ground. The application of copyright to factual data and metadata has even less or no eligibility for copyright protection.

Unfortunately there is no international agreement/definition of copyright, which means that copyright protection is subject to national law. The variety of limitations and exceptions for the users of copyrighted material, which differ per country, is high¹⁰.

However, all European countries and the vast majority of countries worldwide have ratified the Berne Convention, which requires its signatories to treat the copyright of works of authors from other signatory countries at least as well as those of its own nationals¹¹.

The Berne convention prescribes a set of principles of which the following are the most relevant for this report. At first, Article 2 of the Berne Convention states the principle of creativity: only works that are intellectual creations of their authors are eligible for copyright protection. Secondly, only the “*mode or form of expression*” are protected, not the ideas and concepts behind this publication (Art 2 (1), Berne Convention). Thirdly, the work must be fixed on a certain tangible medium (Art 2 (2), Berne Convention). Lastly, Art 5(2) of the Berne Convention states that copyright protection is automatic and does not require formal registration. However, these principles and particularly the principle of creativity are, despite the international nature of this Convention, interpreted differently in various jurisdictions.

A problem of the copyright framework is that its basic concepts are founded on the physical and technical limitations from the print age. The definition of e.g. *reproduction* has changed since the print age (Lammerant et al. et al., 2014).

Database protection right

Although copyright protection for research data in many cases lacks legal ground, the *compilation* or arrangement of research data into a database is recognised by a higher potential to meet the criterion of originality. The classification, coding, formats and interpretations of data in a compilation may broadly be presumed to be covered by copyright (GEOSS, 2011). A compilation is a work formed by the collection and assembling of pre-existing materials or of data that are

¹⁰ In Europe, Article 5 of the Directive 2001/29/EC gives an exhaustive list of exceptions that allow for use of copyrighted material without prior authorisation or remuneration.

¹¹ http://www.wipo.int/treaties/en/text.jsp?file_id=283698

selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship (Chisum *et al.*, 2011).

An alternative or complementary form of a database protection right is the *sui generis* database right, installed for databases that do not meet the criterion of originality for copyright protection. The 1996 Directive on the legal protection of databases (Database Directive), which has been enacted in all EU Member States and Participating States, imposes this statutory form of exclusive property rights protection of databases. This principle is specifically intended to provide for a right for the maker of a “database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilisation of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database” (art 7 (1), Official Journal L 077, 27/03/1996 P. 0020 - 0028). Only the cost for obtaining, creating, or updating the individual data items can be taken into account when assessing the substantial investment. The idea behind this *sui generis* principle is based on the assumption that property rights (i.e. temporary monopoly rights) attract investment and stimulate the economy. This principle distinguishes itself since the copyright laws of many Member States specifically exclude effort and labour from the criteria for copyright protection.

National Security Laws

In some countries national security laws can withhold or restrict the use of data. These laws forbid to make use of the protected data and to disseminate the data or the information based on these data to the public. Such restriction can be found in the field of military defence but also to fight criminal behaviour such as illegal fishing.

Confidentiality Laws and Policies

Confidentiality laws can also protect public and private interests to withhold certain data and information. This confidentiality may consider the public (administrative consultations) as well as the private sector (GEOSS, 2015). An example is sharing the fish catch data of a vessel; the enterprise might want to withhold this information out of commercial interest, to protect its fishing strategy.

Protection of Endangered Species

Data and information on endangered species must, in certain circumstances, be withheld in the interest of protection of these species. An example is the publication of data concerning the precise location of sedentary fish species, publishing this data would jeopardise the existence of these sedentary fish species.

Individual Contracts or Use Agreements

Individual contracts and agreements are only binding to the parties who have signed these arrangements. Once signed, contractors and parties can impose restrictions to the access and reuse of research data and information, which might go far beyond legal justified interests such as described above. The individual contracts are often referred to as end-user licenses agreements (EULAs).

Privacy

Personal data can, under EU law, only be gathered legally under strict conditions, i.e. for a legitimate purpose. Furthermore, persons or organisations that collect and manage personal information must protect it from misuse and must respect certain rights of the data owners, which are guaranteed by EU law. Data is considered personal when it enables anyone to link information to a specific person, even if the person or entity holding that data cannot make that link (Directive 95/46/EC). In the fishery and marine sciences sector, data characterised by a high resolution that e.g. shows the exact location of a vessel at a moment in time is protected through privacy regulation. However, this is according to interviews also sometimes misused as a mere ‘excuse’ not sharing data.

3.1.2 Legal interoperability instruments

A variety of laws and regulations restrict the access, use and reuse¹² of data based on the aforementioned legitimate interests. When data from multiple sources are combined, the new combined database will be restricted by the accumulation of rights from the sources used. Therefore it is of the utmost importance to adopt legal instruments that ensure the compatibility of legal rights, terms, and conditions of databases, i.e. legal interoperability instruments.

According to the Group on Earth Observations (GEO), legal interoperability among multiple datasets from different sources occurs when (2012):

- use conditions are clearly and readily determinable for each of the datasets,
- the legal use conditions imposed on each dataset allow creation and use of combined or derivative products, and
- users may legally access and use each dataset without seeking authorisation from data creators on a case-by-case basis, assuming that the accumulated conditions of use for each and all of the datasets are met.

Legal interoperability can be implemented through a top-down or bottom-up approach. A top-down approach requires the establishment of a global agency, e.g. the United Nations or any of the UN specialised agencies. A bottom-up process to achieve legal interoperability must be based on a step-by-step model that consults all relevant parties. A bottom-up approach would require multi-stakeholder coordination; this approach is therefore more time-consuming but is expected to be more successful than a top-down approach.

As well as a need for two approaches, there is a need for two types of legal interoperability. On the one hand, there is a need for international legal interoperability (such as harmonisation); on the other hand, there is a need for legal interoperability between data providers (such as licensing).

¹² The PSI Directive (Directive 2003/98/EC) defines ‘re-use’ in Article 2(4) as “the use by persons or legal entities of documents held by public sector bodies, for commercial or non-commercial purposes other than the initial purpose within the public task for which the documents were produced. Exchange of documents between public sector bodies purely in pursuit of their public tasks does not constitute re-use.”

This document decided to include standardisation, harmonisation and mutual recognition as relevant *international legal interoperability instruments*. Waivers, common-use licenses, contracts and terms of use are discussed as relevant *legal interoperability instruments*.

3.1.2.1 *International legal interoperability instruments*

Standardisation

Standardisation is usually defined as a regulatory approach that is based on widely accepted good principles, practices or guidelines in a given area; standards may also relate to the usual behaviour of the “reasonable man” (Miller, 2007). A recognised body (such as EGI) should provide common and repeated application of standards on data sharing in the form of rules or guidelines. Standards often qualify as ‘soft law’, i.e. without a legal ground, but can still provide a relevant benchmark for the behaviour of the concerned community. Standardisation can be seen as a first step to harmonisation.

Harmonisation

Harmonisation is the next step in the direction of legal convergence than standardisation. Regulatory harmonisation can be defined as the legal model for institutionalising a desired cooperation by confining actors and policies into the normative standard of rights and obligations (Weber, 2009). Harmonisation seeks to ‘effect an approximation or co-ordination of different legal provision or systems by eliminating major differences and creating minimum requirements or standards’ (Cruz, 1999).

Harmonisation thus targets the coordination between different jurisdictions, so that the same rules will apply to institutes that operate in more than one jurisdiction. When two parties in different jurisdictions want to share data, it often happens the legal systems are too different to accommodate a legal framework (e.g. criteria for copyright protection). Harmonisation of law creates consistency and reduces the complexity of laws, regulations, standards and practices fostering international legal interoperability. However, harmonisation of law is very time-consuming since all rules have to be harmonised individually.

Mutual recognition

Mutual recognition is the consent to compromise a country’s regulatory autonomy, by accepting that another state’s regulation is satisfactory. This means that mutual recognition assumes that different national requirements can be interchangeable in order to apply domestically (Weber, 2009). Mutual recognition is less time-consuming than harmonisation, but is challenging to implement with many countries involved.

3.1.2.2 *Legal interoperability instruments*

Waivers

This is the most favourable written statement for a data user. A rights holder states that no statutory or other rights are retained by this individual, i.e. waives all rights thus data obtains a status equivalent to the public domain.

Common-use Licenses

A license is a unilateral permission by the right holder from the licensor to the licensee to use certain rights. As opposed to waivers, where the rights holder waives all rights, the licensor wishes to retain some rights and control the use of the data. Common-use licenses (further referred to as 'license') distinguish themselves from contracts since the implementation of a license does not require mutual agreement, and is based on existing statutory rights for the implementation (White Paper, GEOSS). A common use license is a unilateral declaration by a copyright owner that he or she allows the use of his or her copyrighted work under the conditions of the specified license. It is not a contract because there is no acceptance and therefore is not enforceable in itself. It can be enforceable on the basis of copyright law, if a certain use represents a copyright infringement.

The main question of licensing is whether the license remains in the IPR regime or whether it becomes a contract. When the license departs from default law and includes new obligations, it becomes a contract (Lammerant et al., 2014). A contract is only binding for parties involved in this agreement, i.e. only for parties that accepted this contract.

When a license is violated, the licensor has to rely on contract law which is much less harmonised than IPR. This creates difficulties when licenses are subject to multiple jurisdictions (Lammerant et al., 2014).

The selection of a license for data depends on many aspects such as the legal suitability for the type of shared data. The most commonly used type of license, which are general purpose licensing tools, is developed by the Creative Commons (CC) Corporation. The CC Corporation is a "non-profit organisation devoted to expanding the range of creative works available for others to build upon legally and to share". The CC licenses have been reviewed in over 70 countries and are in use throughout the world.

The CC licenses offer to right holders a menu of elements from which they can pick their favourite combination: Attribution (BY), Non-Commercial (NC), No Derivative Works (ND, only verbatim copies are allowed), and Share Alike (SA, author requires the creators of derivative works to adopt identical license). The combination of these four elements results in six different licenses:

| |
|--|
| Attribution (BY) |
| Attribution No Derivatives (BY-ND) |
| Attribution Non-Commercial (BY-NC) |
| Attribution Non-Commercial No Derivatives (BY-NC-ND) |
| Attribution Non-Commercial Share Alike (BY-NC-SA) |
| Attribution Share Alike (BY-SA) |

The CC Corporation also developed a rights waiver or transfer to the public domain (i.e. the copyright owner waiving all its rights, including the database right and the right to be identified as the creator): CC Zero (CC0). The CC0 provides an irrevocable, royalty-free and unconditional license for anyone to use the resource for any purpose. Although CC0 doesn't legally require users of the data to cite the source, it does not take away the moral responsibility to give attribution, as

is common in scientific research. This type of license is relatively easy and straightforward to implement.

There are many benefits of using the CC licenses such as the ease of use of the licences, widespread adoption of the licences, and their familiarity and flexibility. The CC licences are also available in both human-readable and machine-readable forms.

However, there are some considerations before adopting the use of CC licenses. A CC licence cannot be revoked once it has been issued. This means that the choice for a CC license should form part of the overall direction of the data provider and most probably will be a high-level policy decision. Many CC licenses also have proven to lack interoperability with themselves. For example, whilst the CC BY-SA licence is more open than the CC BY-ND, it is less interoperable (see table below). Similarly, CC BY-NC-SA and CC-BY-SA licensed data can only be used with themselves (and not with each other), with CC BY licensed data (or equivalent licensed data), or with data released under CC0 (Korn & Oppenheim, 2011). Important to note here is that the CC Corporation has just extended the compatibility with licenses outside the CC license family to the CC BY-NC-SA 4.0 licenses.

In the past six years, the CC Corporation has worked with volunteers around the world to translate and adapt the CC licenses to more than 60 jurisdictions. Version 4.0 of the CC license was published in 2013, which updated among other aspects the inclusion of the *sui generis* database protection right within the CC licensing scope and require compliance with the license conditions when applicable to the particular usage, unless explicitly excluded by the licensor¹³. The table below explains all types of CC licenses and answers the most relevant questions.

| License Type | Who can use the resource and under what terms? | Can the licensed data be modified? | Suitability for data, datasets and databases |
|------------------------------------|--|---|---|
| Attribution (CC-BY) | Anyone | Yes, but you must attribute. You must also ensure that you do not impose any restrictions on the whole of the work licensed beyond the terms of this licence. | Not specifically geared towards data, datasets and databases, but can be used with minimal amounts of data (to avoid attribution stacking) and as long as only an “insubstantial” amount of any databases or datasets are reused. |
| Attribution Share Alike (BY-SA) | Anyone | Yes, but you must attribute and if you use or reuse the data etc., you must use the CC BY-SA end user licence for onward licensing | As above. Share Alike 3.0 requirement can impact negatively on interoperability of data and prevent linked open data. |
| Attribution Non-Commercial (BY-NC) | Anyone – for non-commercial purposes only | Yes, but you must attribute. | As above. Although NC restriction does not pose immediate problems, but ambiguity of what constitutes non-commercial may be problematic. There may also |

¹³ <https://creativecommons.org/Version4/>

| | | | |
|--|--|---|---|
| | | | be interoperability problems with linking to data licensed under more permissive terms. |
| Attribution No Derivatives (BY-ND) | Anyone | No, and you must attribute. | As above. Reuse and repurposing of data, datasets and databases not permitted. |
| Attribution Non-Commercial Share Alike (BY-NC-SA) | Anyone, for non-commercial purposes only | Yes, but you must attribute and if you use or reuse the data etc., you must use the CC BY-SA end user licence for onward licensing. | As above. Share Alike 3.0 requirement can impact negatively on interoperability of data and prevent linked open data. Although NC restriction does not pose immediate problems, but ambiguity of what constitutes non-commercial may be problematic. There may also be interoperability problems with linking to data licensed under more permissive terms. |
| Attribution Non-Commercial No Derivatives (BY-NC-ND) | Anyone, for non-commercial purposes only | No and you must attribute | As above. Reuse and repurposing of data, datasets and databases not permitted. Although NC restriction does not pose immediate problems, but ambiguity of what constitutes non-commercial may be problematic. There may also be interoperability problems with linking to data licensed under more permissive terms |
| CC Zero | Anyone | Yes, with no restrictions whatsoever. | Ideal. |

 Table 1: CC licenses (*adapted from Korn & Oppenheim, 2011*)

Contracts

A contract is, unlike a license, only legally binding once all parties involved have expressed agreement. Contracts require *offer, acceptance, intention to create legal relations*, and *consideration*. Concerning the access to databases, formal offer and acceptance are made typically by clicking on agreements online. Contracts are only legally binding for the agreeing parties; the datasets are therefore more prone to the negative effects of leakage. Another drawback of contracts is its complexity, contracts are after all not standard, i.e. the freedom of contract. However, a useful application of contracts may also be which licenses to apply on a dataset or combination of datasets.

Terms of use

In order to facilitate the reuse of data, it is imperative that others know the terms of use for the database and the data content. The terms of use are rules that one must obey in order to use the data or service. The terms of use agreement is mainly used for legal purposes by data providers and databases that store data. A legitimate term of use agreement is legally binding and may be subject to change. A term of use agreement typically includes the following sections: definitions of key words and phrases, user rights and responsibilities (e.g. proper or expected use, privacy policy and accountability) and a disclaimer limiting the liability of the data provider for damages incurred by the user.

3.2 Data structure and size

There are three broad types of data within the fishery and marine sciences sector:

- 1) Primary/raw data: data coming directly from the source;
- 2) Aggregated data/statistics: data processed from primary/raw data;
- 3) And metadata: reference data describing either the primary/raw data or aggregated data/statistics.

In any data project, there are likely to be two components. The first is the data collected, assembled, or generated, i.e. the raw content in the system (e.g. fish catch from a specific vessel). The second component is the data system in which the data is stored and managed.

We usually do not think of data content separate from the system in which it is stored, but the distinction is important in terms of IPRs. The question is what is protected by copyright¹⁴. Factual data has no copyright protection since it doesn't meet the criterion of creativity; it is after all not possible to copyright facts.

A database, on the other hand, can have a thin layer of copyright protection. Deciding what data needs to be included in a database, how to organise the data, and how to relate different data elements are all creative decisions that may receive copyright protection. But the data itself, as mentioned before, is mostly characterised as objective factual data that does not meet the creativity requirement. The *sui generis* principle, which has been adopted in the EU, is explained in chapter 4.1.1 and offers a complementary or alternative database copyright protection.

Because of the different copyright status of databases and data content, different mechanisms are required to manage each. Copyright and Terms of Use can govern the use of databases and some data content (that which is itself original), but contract law, trademarks, and other mechanisms are required to regulate factual data.

It is worth to consider that the system in which data content is stored may be operated by a different actor and offered through specific terms of use. In this case three types of licenses may be involved: (i) the one agreed between the system operator and the primary data owner, (ii) the one selected for derivative product that may differ from the one associated with primary data, and (iii) the one agreed between the system operator and the data consumer.

All these licenses have to be captured by the "terms of use" of the system, i.e., they are part of the rules a consumer must agree to accept when using the system.

¹⁴ <http://data.research.cornell.edu/content/intellectual-property>

As explained in 4.1 evaluating the interoperability among licenses to promote the re-use of data is by far not an easy task. A re-use license concerns at least attribution, copyleft requirement, and control on commercial exploitation of the dataset. Identifying, managing, and applying some forms of control on access and, especially, re-use conditions in heterogeneous contexts is one of the most relevant yet unsolved issues. Despite few languages to specify rights exist, e.g., MPEG-21 REL and Open Digital Rights Language, meeting the aim to formally specify rights is unrealistic unless specific controlled vocabularies are introduced and agreed. The CC Rights Expression Language, for example, goes in this direction by introducing precise definition of its terms, as expressed by the CC licenses (Assante et al, 2015 and 2016).

Data size within the fishery and marine sciences sector is expected to increase profoundly in future years. Technological innovation, such as satellite information to monitor and detect illegal fishing activities, requires big data collection, management and processing. This will form another practical challenge to share data and will require the availability and use of the necessary technical infrastructure and skills for exploiting large amounts of data for their transformation into useful services and decision-making tools, among others.

There are different rules for public sector information¹⁵ (PSI) in the EU. The Directive on the re-use of public sector information (Directive 2003/98/EC, known as the 'PSI Directive') entered into force on 31 December 2003. It focuses on the economic aspects of re-use of information rather than on the access of citizens to information. It encourages the Member States to make as much information available for re-use as possible. It addresses material held by public sector bodies in the Member States, at national, regional and local levels, such as ministries, state agencies, municipalities, as well as organisations funded for the most part by or under the control of public authorities (e.g. meteorological institutes). But it excludes content held by museums, libraries and archives, which is only re-useable if it is made available by the institutions for re-use.

3.3 Global initiatives

There are many initiatives involved in the production, management and sharing of data as well as in encouraging and supporting institutions to share data. This analysis focuses on the fishery and marine sciences sector. At the same time, other global initiatives relevant for their approach to legal interoperability are also discussed. Interesting initiatives with sharing data as core activity and thereby stimulate (legal) interoperability of global data sharing are stated in the non-exhaustive list below.

iMarine

iMarine¹⁶ is the Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources. The final goal of iMarine is to launch an initiative aimed at establishing and operating an e-infrastructure supporting the principles of the Ecosystem Approach to fisheries management and conservation of marine living resources.

¹⁵ Public sector information (PSI) is information produced by central and local government or any other public body.

¹⁶ <http://i-marine.eu>

Adapted to the marine context, the Ecosystem Approach to Fisheries (EAF) has the specific purpose to plan, develop, and manage fisheries in a manner that addresses the multiple needs and desires of societies, without jeopardising the options for future generations to benefit from the full range of goods and services provided by marine ecosystems.

The EAF covers all socioeconomic and environmental aspects of policy-making, fishery assessments, fishing operations, processing and trade, involving a number of interest groups each one having its own societal structure, objectives and economic and computational resources to achieve them. The fishery and environmental scientists interact quite often, use common concepts, approaches and methods and are developing some consensus about the state of the fisheries and their environment, the problems and the solutions potentially available to resolve them.

iMarine is compiled of four ‘cubes’, i.e. the BioCube (generate new knowledge from data), the ConnectCube (collaborative, standards-oriented data publication environment, including semantic technologies), GeoCube (collaborative, standards-oriented data publication environment, including semantic technologies) and the StatsCube (data-life-cycle supporting framework).

Virtual Research Environments (VREs) are created by selecting and combining several iMarine applications to support diverse communities of practice. VREs offer flexible and secure web-based, community-centric platforms, so researchers can work together on common challenges. Each VRE in the infrastructure is tightly integrated with the underlying gCube enabling software, and can access and re-purpose data from other iMarine applications.

iMarine’s unique selling points are that it is a domain specific (thematically focused) rather than a generic approach, with a strong focus on marine environment. Another key benefit is that there is no need for users to invest in hardware and software as they are already part of the infrastructure. The VRE concept and social facilities have also proven to be a good selling point.

iMarine also managed to raise awareness on the importance of adopting a standard exchange protocol for the fisheries operations data domain. The idea of this Fisheries Language for Universal eXchange (Flux) initiative is to streamline data-flows. Chapter six discusses an interview with Marc Taconet, the iMarine Board Chair.

Global Biodiversity Information Facility (GBIF)

GBIF is an international open data infrastructure, funded by governments. It allows anyone, anywhere to access data about all types of life on Earth, shared across national boundaries via the Internet. GBIF strongly encourages data publishers to opt for the most open choice available (CC0).

GBIF is an interesting initiative to analyse since biodiversity data is in essence multidisciplinary. Therefore only broader cooperative frameworks can provide common standards, legitimate interests for the restriction of data and a legal framework for data sharing.

Copyright only applies to creative content and 99% of GBIFs data are facts, which cannot be copyrighted. GBIF does hold copyright over some text in remarks fields, the data format or database model GBIF created.

Figuring out where the facts stop and where the (copyrightable) creative content begins can already be difficult for the content owner, let alone the content user. On top of that different rules are used in different countries. GBIF decided to publish their data under CC0, this removes any ambiguity and red tape. Any copyright protection has thus been waived and the data is placed in the public domain¹⁷.

Major data publishers for GBIF and of interest to this analysis are Fishbase, Ocean Biographic Information System (OBIS), Encyclopedia of Life (EOL) and the Pan-European Species-directories Infrastructure (PESI).

Group on earth Observations (GEO)

A central part of GEO's mission is to build the Global Earth Observation System of Systems (GEOSS). GEOSS is a set of coordinated, independent Earth observation, information and processing systems that interact and provide access to diverse information for a broad range of users in both public and private sectors. It aims to construct a global public infrastructure for Earth observations consisting in a flexible and distributed network of systems and content providers.

The distinctive aspect of this system is that GEOSS links earth observation systems from 96 countries and the EU. GEOSS thus created an initiative where a high number of data providers place their data. GEOSS therefore serves as a good example of (legal) interoperability.

Research Data Alliance (RDA)

RDA enables data to be shared across barriers through focused Working Groups and Interest Groups, formed of experts from around the world – from academia, industry and government.

Participation in RDA is open to anyone who agrees to its guiding principles of openness, consensus, balance, harmonisation, with a community driven and non-profit approach. It was started in 2013 by a core group of interested agencies – the European Commission, the US National Science Foundation and National Institute of Standards and Technology, and the Australian Government's Department of Innovation. Other agencies, countries, companies, associations and institutes are also invited to join. RDA has a broad, committed membership of individuals – now almost 3,000 from 102 countries since RDA was launched in March 2013 - dedicated to improving data exchange.

One of RDA's Working Groups (WGs) is the RDA-CODATA Working Group on Legal Interoperability of Research Data (RDA-CODATA WG). This Working Group chaired by Paul F. Uhler (please see chapter six for an interview with Mr. Uhler), Enrique Alonso Garcia, Bob Chen has done extensive research on legal interoperability of data sharing. The main objective of this WG is to formulate core principles and guidelines of best practices through which legal interoperability can be achieved¹⁸.

iMarine is also represented in RDA through Donatella Castelli who is both a member of the REDA Europe Forum & the Marine Data Harmonisation Interest Group.

AGINFRA

¹⁷ <http://www.canadensys.net/2012/why-we-should-publish-our-data-under-cc0>

¹⁸ <https://rd-alliance.org/about.html>

AGINFRA started as an FP7 project¹⁹, co-funded by the European Commission and it is currently evolving into a global hub for collecting and sharing information related to agriculture & food security, plays multiple roles in the global agri-food research community - including but not limited to the following:

1. a global atlas of agricultural research & extension (including institutions, people, publications, data sets, projects, courses, OERs);
2. a semantic layer of processing, enriching & interlinking research information from distributed, heterogeneous sources & formats;
3. a catalogue of software components (open source software stack & APIs) that anyone may use to process research information;
4. a help desk service to support institutions & projects that wish to publish their research information openly;
5. a set of data-rich service and application demonstrators for specific case studies (such as food safety, viticulture research, crop composition etc.);

Among others, AGINFRA provides the e-infrastructure required by researchers in the agri-food sector so that they have access to the platforms, tools and services required for their research.

CIARD RING

CIARD (www.ciard.info) is a global movement for open agricultural knowledge for development, currently consisting of more than 6,200 members; about 440 of which are institutions. The network is working on the advocacy on open knowledge for agricultural development, promoting open access to agricultural knowledge. CIARD is working mostly on capacity development and in this context, it produces capacity building material in the form of pathways, webinars, e-discussions, working groups and an advocacy toolkit²⁰, all of which aim at facilitating access to agricultural research outcomes so that they become available to all types of stakeholders. In addition, CIARD has published the “CIARD Manifesto Towards a Knowledge Commons on Agricultural Research for Development”²¹ that supports adoption of the coherent, effective and open institutional approaches to agricultural knowledge.

CIARD is also responsible for and maintaining the RING²², a global directory of web-based services that provide access to any kind of information sources pertaining to agricultural research for development (ARD). CIARD RING provides access to resources such as providers, services and datasets. CIARD RING is one of the core components of the AGINFRA global agri-food research e-infrastructure.

GODAN

GODAN²³, the Global Open Data for Agriculture and Nutrition is a global network that currently features more than 200 members, including a variety of stakeholder types, such as national

¹⁹ <http://www.aginfra.eu/project/>

²⁰ <http://www.ciard.info/capacity-development>

²¹ www.ciard.info/about/manifesto

²² <http://ring.ciard.info>

²³ www.godan.info

governments, non-governmental organisations, international and private sector organisations. The aim of GODAN is to:

- 1) advocate for new and existing open data initiatives to set a core focus on agriculture and nutrition data;
- 2) encourage the agreement on and release of a common set of agricultural and nutrition data;
- 3) increase awareness of ongoing activities, innovations, and good practices;
- 4) advocate for collaborative efforts on future agriculture and nutrition open data endeavours;
- 5) advocate programs, good practices, and lessons learned that enable the use of open data particularly by and for the rural and urban poor.

Among the key partners and supporters of GODAN are the US Government, the UK Department for International Development (DFID), the Netherlands Government, the Open Data Institute (ODI), the Food and Agriculture Organisation of the United Nations (UN FAO), the Technical Centre for Agricultural and Rural Cooperation ACP-EU (CTA), the Centre for Agriculture and Biosciences International (CABI), CGIAR (the ex-Consultative Group for International Agricultural Research) and the Global Forum on Agricultural Research (GFAR).

One of the latest outcomes of GODAN is the Discussion Paper on Open Data in Agriculture & Nutrition discussion paper, jointly developed by GODAN and the Open Data Institute (ODI) in May 2015. This publication highlights the role of open data in the context of agriculture and nutrition, identifies a number of challenges and presents several use cases where open data made an impact in the agri-food sector.

Of course this list is non-exhaustive; other important global initiatives are the Taxonomic Database Working Group (TDWG), Fishbase, Fridjoff Nansen programme, Tuna Atlas, Vulnerable Marine Ecosystems database (VME) and Aquatic Sciences and Fisheries Abstracts (ASFA). However, due to limitations in terms of time and size of this report, only an indicative set of data stakeholders in the fishery and marine sectors is presented and analysed.

Other International Organisations also published documents on the legal aspects of data sharing. The Organisation for Economic Cooperation and Development (OECD) is leading on this aspect, and has developed the Principles and Guidelines for Access to Research Data from Public Funding. These principles and guidelines are meant to apply to research data that are gathered using public funds for the purposes of producing publicly accessible knowledge.

3.4 EU initiatives

Interesting initiatives with sharing data as core activity and thereby stimulate (legal) interoperability of data sharing in the EU are stated in the non-exhaustive list below.

EGI-Engage

The EGI-Engage project (Engaging the Research Community towards an Open Science Commons) started in March 2015, co-funded by the European Commission for 30 months, as a collaborative effort involving more than 70 institutions such as the FAO and CERN in over 30 countries. EGI-Engage aims to accelerate the implementation of the Open Science Commons by expanding the capabilities of a European backbone of federated services for compute, storage, data,

communication, knowledge and expertise, complementing community-specific capabilities. EGI-Engage does not host data under its direct control but functions as a service provider for research communities. The biggest community of users is hosted within CERN. EGI-Engage hosts four major experiments: the ATLAS Data Access Policy, the ALICE data preservation strategy, the CMS data preservation, reuse and open access policy and the LHCb External Data Access Policy. These four experiments all agree with open access to data (CC0 waiver), software and documentation that will allow the processing of data by external entities under conditions formulated by the initiatives themselves with support of EGI-Engage.

EGI-Engage's Open Science Commons is grounded on three pillars:

- 1) the e-Infrastructure Commons, an ecosystem of key services;
- 2) the Open Data Commons, where any researcher can access, use and reuse data;
- 3) and the Knowledge Commons, in which communities have shared ownership of knowledge and participate in the co-development of software and are technically supported to exploit state-of-the-art digital services.

EGI-Engage aims to expand the capabilities offered to scientists (e.g. improved cloud or data services) and the spectrum of its user base by engaging with large Research Infrastructures (RIs), the long-tail of science and industry or SMEs. The main engagement instrument will be a network of eight Competence Centres, where National Grid Initiatives (NGIs), user communities, technology and service providers will join forces to collect requirements, integrate community-specific applications into state-of-the-art services, foster interoperability across e-Infrastructures, and evolve services through a user-centric development model. The project will also coordinate the NGI efforts to support the long-tail of science by developing ad hoc access policies and by providing services and resources that will lower barriers and learning curves²⁴.

Infrastructure for Spatial Information in the European Community (INSPIRE)

INSPIRE²⁵ serves as a useful example for the development of technical interoperability. INSPIRE is "an EU initiative to establish an infrastructure for spatial information in Europe that is geared to help to make spatial or geographical information more accessible and interoperable for a wide range of purposes supporting sustainable development".

The unique aspect of this initiative is the requirement of INSPIRE to adopt common Implementing Rules (IR) in a number of specific areas (Metadata, Data Specifications, Network Services, Data and Service Sharing and Monitoring and Reporting). These IRs are adopted as Commission Decisions or Regulations, and are binding in their entirety.

European Bioinformatics Institute (EMBL-EBI)

²⁴ <https://www.egi.eu/about/egi-engage/overview.html>

²⁵ <http://inspire.ec.europa.eu/>

As mentioned before, ‘big data’ will gain relevance in future years also in the fishery and marine sciences sector. The objective of EMBL-EBI²⁶ is to help scientists realise the potential of big data in biology. EMBL-EBI manages the world’s public biological data and makes it freely available to the scientific community via a range of services and tools, perform basic research and provide professional training in bioinformatics.

EMBL-EBI is a pivotal partner in several of Europe’s emerging research infrastructures. It plays a key role in ELIXIR, the emerging infrastructure for biological information in Europe, and BioMedBridges, a project that built technical bridges between data and services in the biological, medical, translational and clinical domains..

International Council for the Exploration of the Sea (ICES)

ICES is an intergovernmental organisation whose main objective is to increase the scientific knowledge of the marine environment and its living resources and to use this knowledge to provide unbiased, non-political advice to competent authorities. All data held by ICES should eventually be placed in the public domain. However, some restrictions such as the use or reuse for commercial purposes and sensitivity of data can be legitimate. Chapter six of this document will discuss an interview held with a representative from ICES.

L’Institut de recherche pour le développement (IRD)

As a French science and technology institute, the IRD is under the joint supervision of the Ministries of Research and Foreign Affairs. This multidisciplinary research institute addresses international development issues. This interdisciplinary approach imposes many legal interoperability issues for IRD, this makes it an interesting institute to analyse. The author of this document had the opportunity to interview a representative from this institute, which will be discussed in chapter six of this document.

LifeWATCH

LifeWATCH²⁷ is a European Research Infrastructure for Biodiversity that serves as an integrated access to a variety of data, analytical and modelling tools as served by a variety of collaborating initiatives. It is working with other international organisations in different areas. Improving the connections and collaboration among the scientific community in order to create a European area of knowledge is one of its objectives.

²⁶ <http://www.ebi.ac.uk/>

²⁷ <http://www.lifewatch.eu>

4 Consultation on Stakeholders and Data Sources

This chapter identifies the key stakeholders in the domain of a legal framework for interoperability. The stakeholders are classified according to their involvement with legal instrumentation: interoperability responsibility type, data managed, type of legal instruments used, activities performed, and overall interest in legal interoperability. The scope of the legal interoperability framework in fisheries is illustrated by a simplified value chain. Several stakeholders were asked to provide details, especially where their framework overlaps with the interest of EGI, e.g. where there are legal requirements surrounding data management. The chapter concludes with a summary of initial insights and preliminary findings.

4.1 Identification of Stakeholders

The domains of interest in fishery and marine science were detailed in Chapter 4 of D2.7 ‘Market Analysis’. This deliverable limited the list to those with a vested interest in legal interoperability:

1. Marine Fisheries Exploitation and Monitoring
2. Fisheries/Aquaculture Catches - Traceability/Certification/Quality Control
3. Marine Fisheries Research
4. Marine Fisheries Policy Making/Management

The relevance of scope of legal interoperability depends on the use of the data, and we recognise three main categories, similar to those of D2.7:

- Monitoring; national public sector organisations monitoring fisheries
- Management; (a) provide fisheries management recommendations and (b) provide plans and measures to national and regional governmental organisations
- Exploitation; mostly companies in the private sector; SME's and industries managing often confidential fisheries operations data

The relevant stakeholders require an active management of legal interoperability; i.e. infrastructure based services that maintain and manage interoperability. Other stakeholders may encounter legal interoperability issues, but without practical engagement with policy or service development to address interoperability, they were excluded from this study. We included:

1. Fisheries management institutions/Authorities
2. Fisheries Research Institute
3. Regional Fisheries Management Organisations
4. Monitoring Control and Surveillance Organisations
5. International Organisations

The 5 stakeholder categories above are not exclusive. Other public or private may also have an interest in legal interoperability, such a legal research centres, but we focus on data managers.

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.

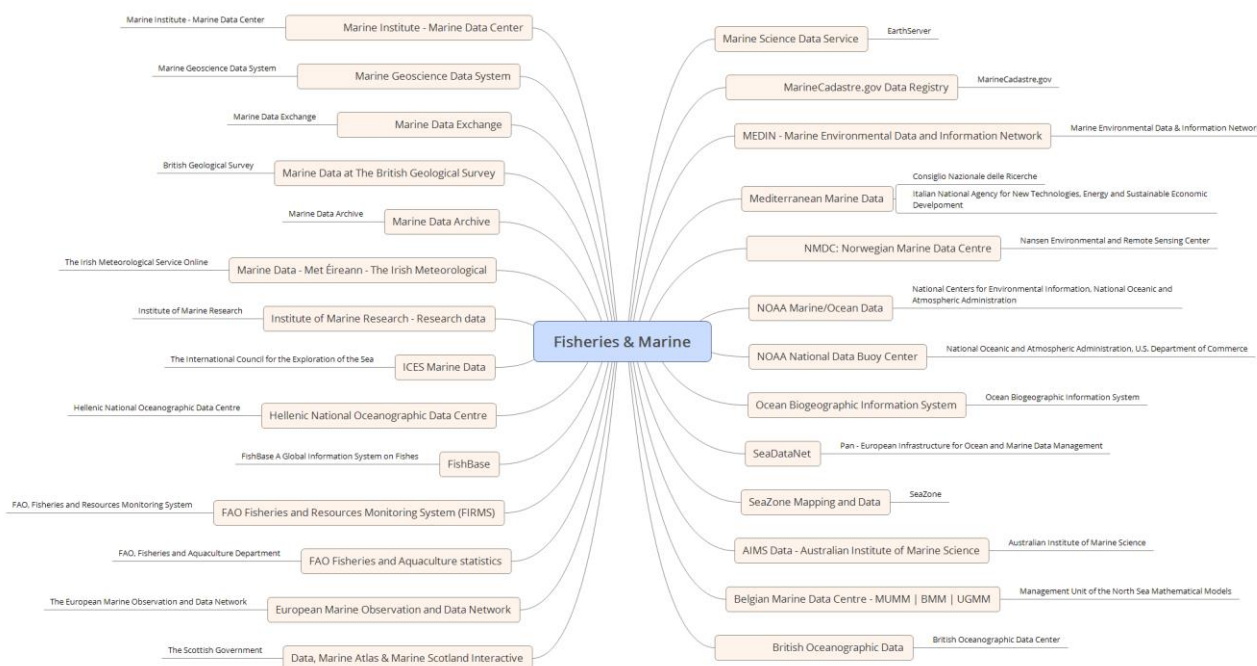


Figure 1: Indicative view of the marine and fisheries key stakeholders

4.2 Stakeholder classification

The stakeholders can be classified by assessing their involvement with Legal Interoperability (LI) definition, instruments, or infrastructure services. Below is a summary list of what their legal interoperability interests (ref. D2.7 for Activity and scope notes).

| 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 Using LI instruments |
| Fisheries Research Institute | | <ul style="list-style-type: none"> Activity - monitor fisheries, provide recommendations for fisheries management Scope of data - national data Using LI instruments, and using LI infrastructure services |
| Regional Fisheries Management Organisation (RFMO) | | <ul style="list-style-type: none"> Activity - provide recommendations on stocks (stock assessment) Scope of data - regional data Using LI instruments, and using some LI infrastructure services |
| Monitoring Control and Surveillance Organisation | | <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 LI definition, using LI instruments |
| International Organisation | | <ul style="list-style-type: none"> Activity - centralise data through definition of data calls and standards, which are shared with state members, fishery management (harmonisation) for EU level with a global impact |

- Scope of data - EU and global view (fleets operating at globally)
- LI definition, using LI instruments, and using LI infrastructure services

4.3 Indicative key data stakeholders

This section lists key stakeholders in the fisheries, marine and maritime sectors that act as data providers. The following figure provides an indicative view of the ecosystem of data stakeholders in marine and fisheries. A number of these data sources are described in the following sections.

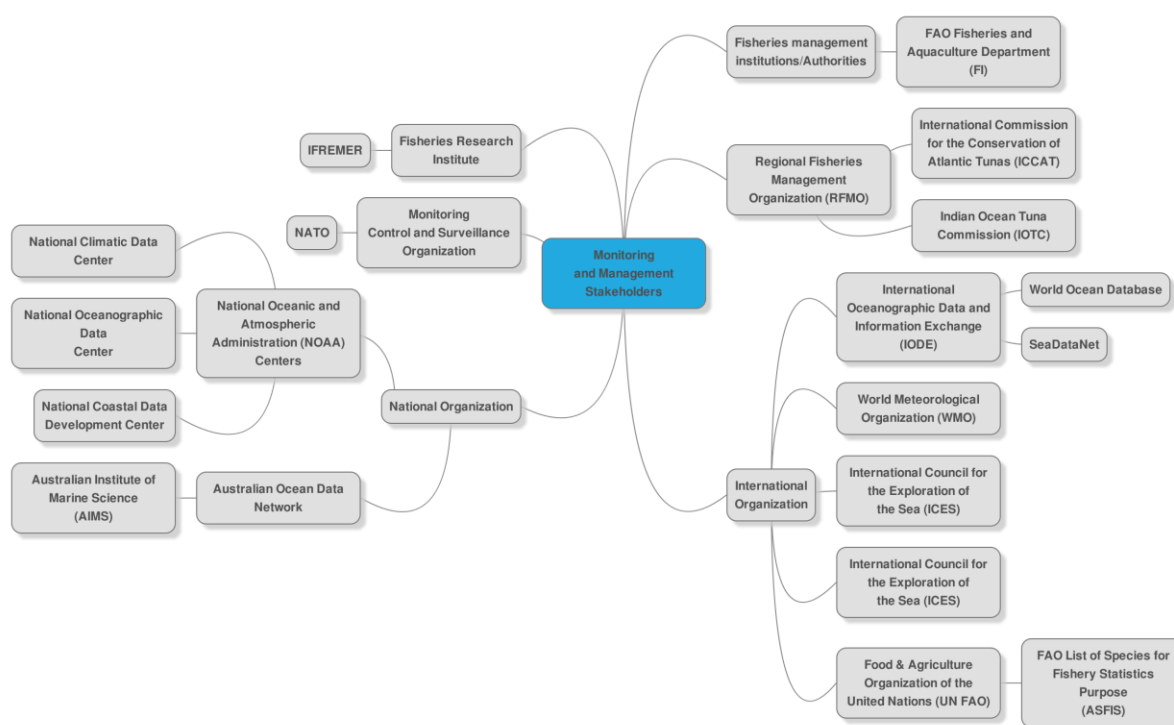


Figure 2: Indicative view of the marine and fisheries data stakeholders' ecosystem

4.3.1 IODE – International Oceanographic Data and Information Exchange

The programme "International Oceanographic Data and Information Exchange" (IODE²⁸) of the "Intergovernmental Oceanographic Commission" (IOC) of UNESCO was established in 1961. Its purpose is to enhance marine research, exploitation and development, by facilitating the exchange of oceanographic data and information between participating Member States, and by meeting the needs of users for data and information products.

²⁸ <http://www.iode.org>

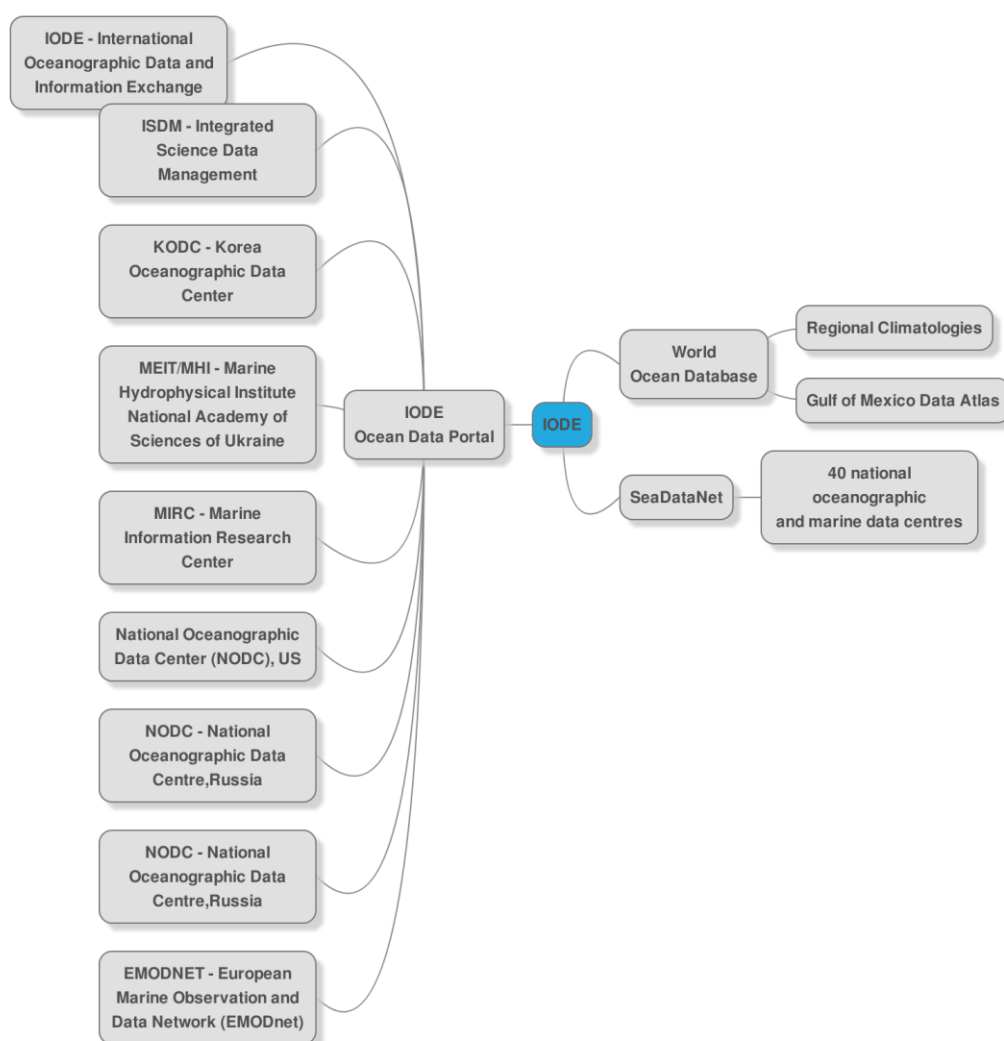


Figure 3: Mapping of the IODE network

Based on the revision IODE-XXII (March 2013), the main objectives of the IODE Programme are:

- To facilitate and promote the discovery, exchange of, and access to, marine data and information including metadata, products and information in real-time, near real time and delayed mode, through the use of international standards, and in compliance with the IOC Oceanographic Data Exchange Policy for the ocean research and observation community and other stakeholders;
- To encourage the long term archival, preservation, documentation, management and services of all marine data, data products, and information;
- To develop or use existing best practices for the discovery, management, exchange of, and access to marine data and information, including international standards, quality control and appropriate information technology;
- To assist Member States to acquire the necessary capacity to manage marine research and observation data and information and become partners in the IODE network;
- To support international scientific and operational marine programmes, including the Framework for Ocean Observing for the benefit of a wide range of users.

IODE provides access to three major networks and data sources, among others:

1. IODE Ocean Data Portal
2. World Ocean Database
3. SeaDataNet

These data sources will be described in the following sections.

4.3.2 National Oceanic and Atmospheric Administration (NOAA) Centers

The National Centers for Environmental Information (NCEI)²⁹ of the National Oceanic and Atmospheric Administration of the U.S.³⁰ are responsible for hosting and providing public access to one of the most significant archives for environmental data on Earth with over 20 petabytes of comprehensive atmospheric, coastal, oceanic, and geophysical data. After some relatively recent structural changes, NOAA's former three data centers (the National Climatic Data Center, the National Geophysical Data Center, and the National Oceanographic Data Center, which includes the National Coastal Data Development Center, respectively) have merged into the National Centers for Environmental Information (NCEI).

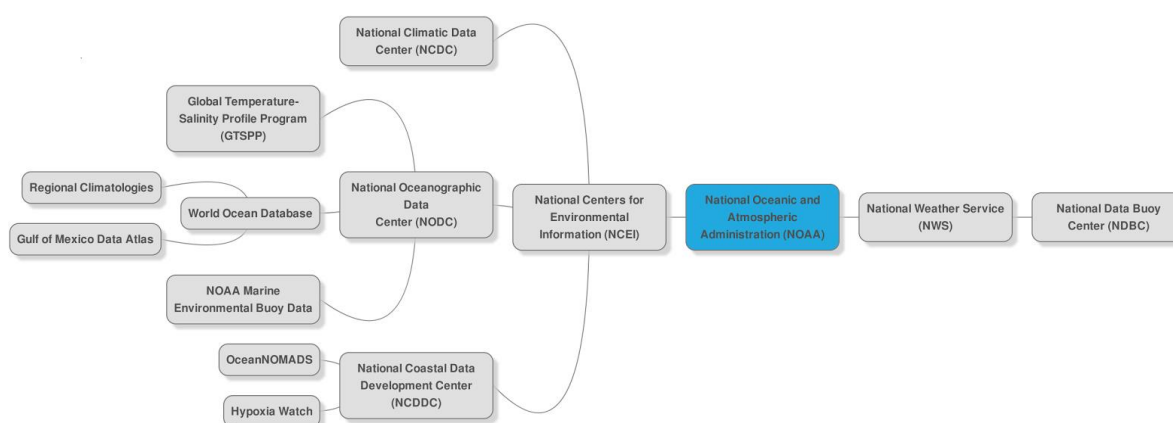


Figure 4: NOAA Centers and data services of interest to this report

NCEI is responsible for hosting and providing access to one of the most significant archives on Earth, with comprehensive oceanic, atmospheric, and geophysical data. From the depths of the ocean to the surface of the sun and from million-year-old sediment records to near real-time satellite images, NCEI is the US leading authority for environmental information.

4.3.2.1 National Climatic Data Center

The National Climatic Data Center (NCDC)³¹ receives and archives meteorological data from ships at sea, moored and drifting buoys, coastal stations, rigs, and platforms. The temporal frequency of the observations range from sub-hourly to six-hourly synoptic and are global in spatial coverage.

²⁹ <https://www.ncei.noaa.gov>

³⁰ <https://www.noaa.gov>

³¹ <https://www.ncdc.noaa.gov>

4.3.2.2 *National Oceanographic Data Center*

Of interest to this analysis is NCEI's National Oceanographic Data Center (NODC)³², which acquires, preserves, monitors, and assesses the U.S. treasure of oceanographic data and information. NODC provides a wealth of data-powered services to its users. NODC maintains two ocean profile databases; the first is the Global Temperature-Salinity Profile Program (GTSP) and the second is the World Ocean Database (WOD).

The GTSP database³³ is comprised of real-time data transmitted over the Global Telecommunication System (GTS) and delayed-mode data from national oceanographic data centers which participate in the International Oceanographic Data and Information Exchange (IODE) system of the Intergovernmental Oceanographic Commission (IOC). Most profiles contain temperature and salinity as a function of depth. The GTSP focuses on managing data collected after 1990. At the request of the oceanographic community, each GTSP profile has an associated history file that contains information about any changes made to the profiles or their associated metadata for errors that were discovered during processing. GTSP is updated on a daily basis and serves the needs of the real-time oceanic and atmospheric forecasting communities.

The World Ocean Database (WOD)³⁴ is the world's largest collection of ocean vertical profile data. The WOD contains profiles dating back to the 19th century and includes data from over 20 variables such as temperature, oxygen, nutrients, plankton, and tracers. Every three months, the WOD includes data from GTSP by applying additional quality control procedures to the latest GTSP collection. Errors or problems discovered through the WOD quality control processes are reported back to GTSP for inclusion in that database and its methods, if applicable. The WOD also includes historical data recovered as part of the IOC/IODE Global Oceanographic Data Archaeology and Rescue project. Substantial amounts of additional data received at NODC and its collocated World Data Center for Oceanography, are also included in the WOD. The WOD contains more detailed quality control flags than GTSP does. The WOD is widely used for ocean climate diagnostic analyses, for long-term ocean data assimilations in numerical models, and for comparisons with satellite data.

Efforts are underway to improve the synchronisation between the GTSP and WOD databases so users can visit the NODC website, and with one selection, acquire all data from both databases with no duplication. Both projects work closely together to mutually improve their databases and processes, which will better serve NODC's user communities.

4.3.2.3 *National Coastal Data Development Center*

The National Coastal Data Development Center (NCDDC)³⁵ provides access to coastal data. The center's activities address NOAA's long-term goals and span across NOAA's Line offices, addressing the scientific stewardship of those data sets. NCDDC provides access to a number of data sources

³² <http://www.nodc.noaa.gov>

³³ <http://www.nodc.noaa.gov/GTSP>

³⁴ <http://www.nodc.noaa.gov/OC5/SELECT/dbsearch/dbsearch.html>

³⁵ <http://www.ncddc.noaa.gov>

and data-powered products, such as OceanNOMADS³⁶ and Hypoxia Watch³⁷ that are described in the following sections.

4.3.3 World Meteorological Organization (WMO)

The World Meteorological Organization (WMO)³⁸ is a specialised UN agency, established in 1951. Its headquarters are in Geneva (Switzerland), and it was the successor to the International Meteorological Organization (IMO), whose origins date back to 1853. The WMO has 185 Member States, and is responsible for global cooperation in meteorological and hydrological observations and services (including systems for rapid data exchange, standardised observations and uniform publication of observations and statistics). The backbone of WMO is the WWW or ‘World Weather Watch’, a global data and information network of measuring stations, managed by Member States and using nine satellites, plus approximately 10 000 land-based, 7 000 ship-based and 300 fixed and floating measuring buoys with automatic weather stations.

Two important services for operational oceanography are provided by WMO:

1. The Global Ocean Observing System (GOOS)³⁹ and
2. Global Climate Observing System (GCOS)⁴⁰

GOOS includes two main groups of operations: (i) measuring systems in the open ocean, specifically to support services at sea, weather prediction, and monitoring of climate change; and (ii) measuring systems in coastal areas, aimed at the study of the health and sustainable development of these areas. GOOS was initially established on the basis of existing observing systems, but it also developed its own pilot projects such as GODAE. On the other hand, GCOS addresses the total climate system including physical, chemical and biological properties, and atmospheric, oceanic, terrestrial, hydrologic, and cryospheric components.

4.3.4 SeaDataNet

The SeaDataNet infrastructure⁴¹ links 40 national oceanographic and marine data centres from 35 countries riparian to all European seas. The data centres manage large sets of marine and ocean data, originating from their own institutes and from other parties in their country, in a variety of data management systems and configurations. Datasets may be browsed or searched using predefined filters. It is one of the three data sources available through IODE.

4.3.5 International Council for the Exploration of the Sea (ICES)

The International Council for the Exploration of the Sea (ICES)⁴² is a global organisation that develops science and advice to support the sustainable use of the oceans. ICES is a network of more than 4000 scientists from over 350 marine institutes in 20 member countries and beyond.

³⁶ <http://ecowatch.ncddc.noaa.gov>

³⁷ <http://www.ncddc.noaa.gov/hypoxia/products>

³⁸ <http://www.wmo.ch>

³⁹ <http://www.ioc-goos.org>

⁴⁰ <http://www.wmo.int/pages/prog/gcos>

⁴¹ <http://www.seadatanet.org/Data-Access>

⁴² <http://ices.dk/Pages/default.aspx>

ICES prioritises, organises, delivers, and disseminates research needed to fill gaps in marine knowledge related to ecological, political, societal, and economic issues.

ICES delivers scientific publications, information and management advice requested by member countries and international organisations and commissions such as the Oslo Paris Commission (OSPAR), the Helsinki Commission - Baltic Marine Environment Protection Commission (HELCOM), the North East Atlantic Fisheries Commission (NEAFC), the North Atlantic Salmon Conservation Organization (NASCO), and the European Commission (EC).

4.3.6 Australian Ocean Data Network

The Australian Ocean Data Network (AODN)⁴³ is an interoperable, online network of marine and coastal data resources, including data from the six (6) Commonwealth agencies with primary responsibility for marine data:

1. The Australian Institute of Marine Science
2. Geoscience Australia
3. Royal Australian Navy
4. Australian Bureau of Meteorology
5. Commonwealth Scientific and Industrial Research Organisation
6. Australian Antarctic Division
7. Integrated Marine Observing System (IMOS)

Since its inception, the AODN has grown to encompass organisations and individual members of the Australian, New Zealand and Pacific marine research community.

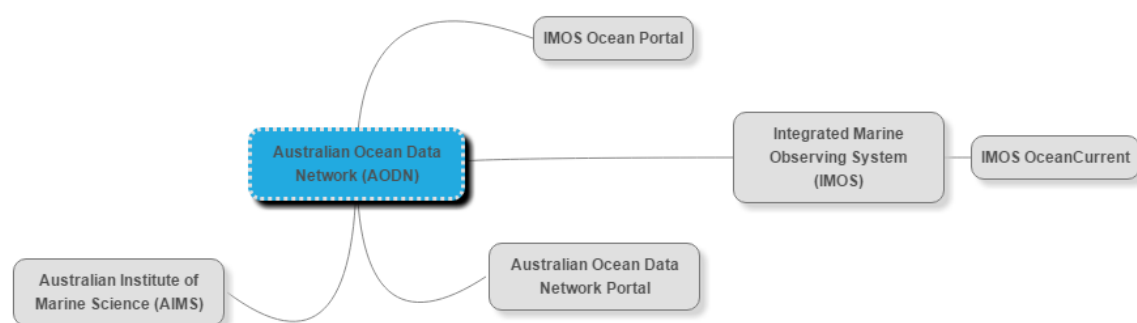


Figure 5: Structure of the Australian Ocean Data Network

The infrastructure of the AODN is based on the IMOS Ocean Portal⁴⁴, which follows international standards and agreements for data/metadata formatting, discovery and sharing.

4.3.7 iMarine

The iMarine Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources⁴⁵ was an open and collaborative initiative aimed at supporting the

⁴³ <http://portal.aodn.org.au/aodn/>

⁴⁴ <https://imos.aodn.org.au>

⁴⁵ <http://i-marine.eu/>

implementation of the Ecosystem Approach to fisheries management and the conservation of living marine resources. More specifically, iMarine provides an e-infrastructure that facilitates open access and the sharing of a multitude of data, collaborative analysis, processing and mining processing, as well as the publication and dissemination of newly generated knowledge. iMarine currently features more than 2,000 users from various scientific domains, such as fisheries, biodiversity and ocean observation, making use of its 15 ready-to-use services. The iMarine e-infrastructure provides access to over a billion records hosted in more than 50 worldwide repositories.

Among others, iMarine features a selected list of data providers of relevance to the marine and fisheries research communities, which is available online⁴⁶.

4.3.8 Eurostat

Eurostat⁴⁷ is the statistical office of the European Union. Its task is to provide the European Union with statistics at European level that enable comparisons between countries and regions. Eurostat offers a whole range of data that governments, businesses, the education sector, journalists and the public can use for their work and daily life. It should be noted that Eurostat does not collect data; this is done in Member States by their statistical authorities, which in turn verify and analyse national data and send them to Eurostat. Eurostat's role is to consolidate the data and ensure they are comparable, using harmonised methodology. Eurostat is actually the only provider of statistics at European level.

Eurostat is another quality source of fisheries and aquaculture statistical data⁴⁸, including data on catches, landings, aquaculture and fleet statistics. Together with the Commission Maritime Affairs and Fisheries DG, Eurostat also keeps data on the number, size, tonnage, power and age of the EU fishing vessels. Data available through Eurostat is visualised and the primary data can also be accessed in the form of spreadsheets that can be downloaded.

4.4 Fisheries and Marine Data Coverage

The EGI Engage Deliverable 2.7; Market Report on the Fishery and Marine Sciences Data Analysis Sector, produced the analysis of the data coverage in the fisheries domain. This resulted in the following graph.

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.

⁴⁶ <http://i-marine.eu/Pages/ShowContent.aspx?id=599ee2b9-de2e-4ab2-80ee-808a9a973b13>

⁴⁷ <http://ec.europa.eu/eurostat>

⁴⁸ http://ec.europa.eu/eurostat/statistics-explained/index.php/Fishery_statistics

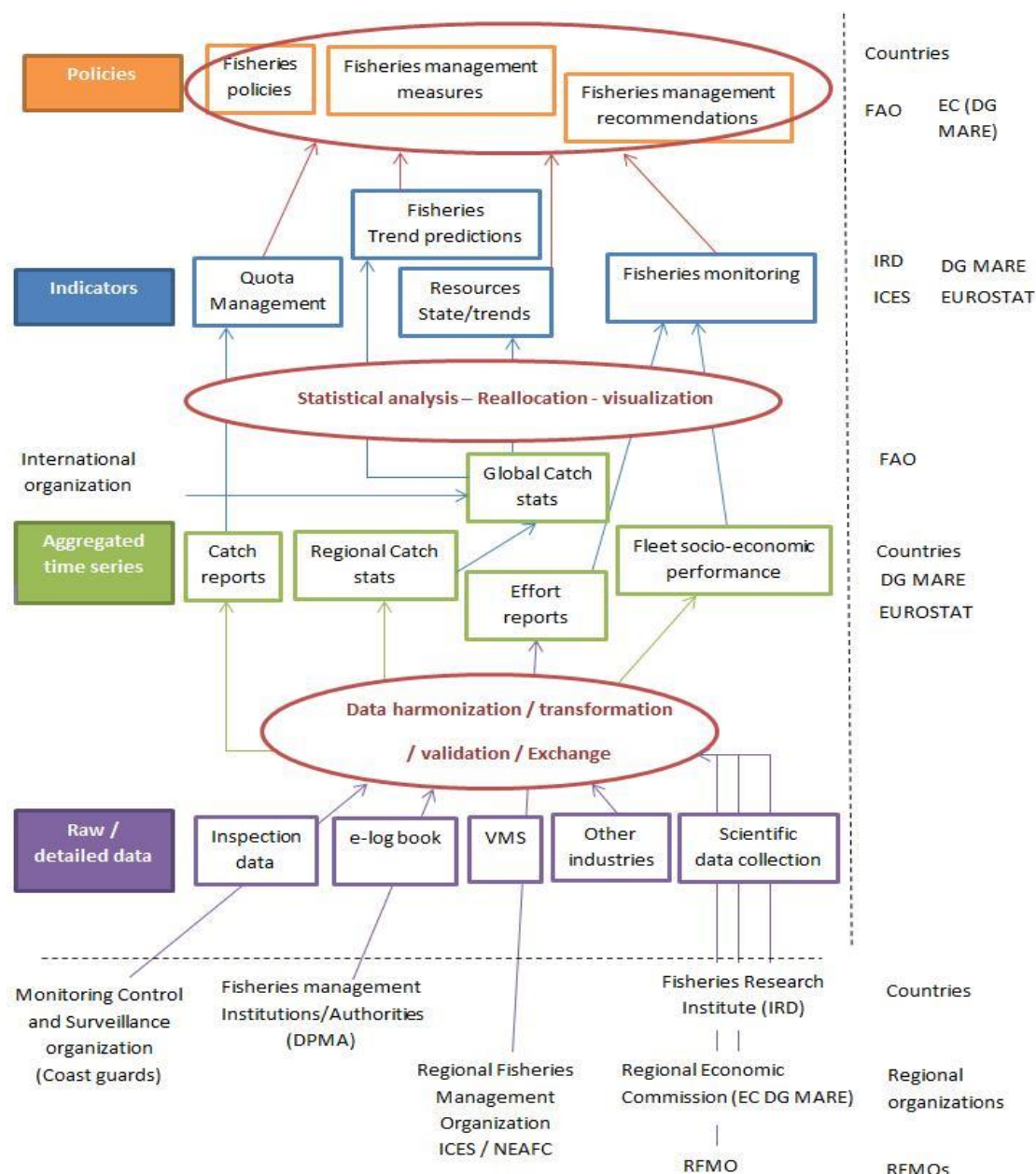


Figure 6: Flow of data and information between various actors

Similar graphs were produced for the aquaculture sector and can be produced for other marine data management sectors, and can be found in D2.7. This deliverable focuses on the fisheries sector, where the legal interoperability aspects are most evident due to the large amounts of systems managing sensitive and confidential data. Therefore, stakeholders from this sector were selected to obtain a relevant representation of most related domains.

4.5 Analysis of Indicative Fisheries and Marine Data Sources

The following sections provide a description of data sources of interest to this analysis. The following figure provides an indicative view of the ecosystem of data stakeholders in marine and fisheries. A number of these data sources are described in the following sections.

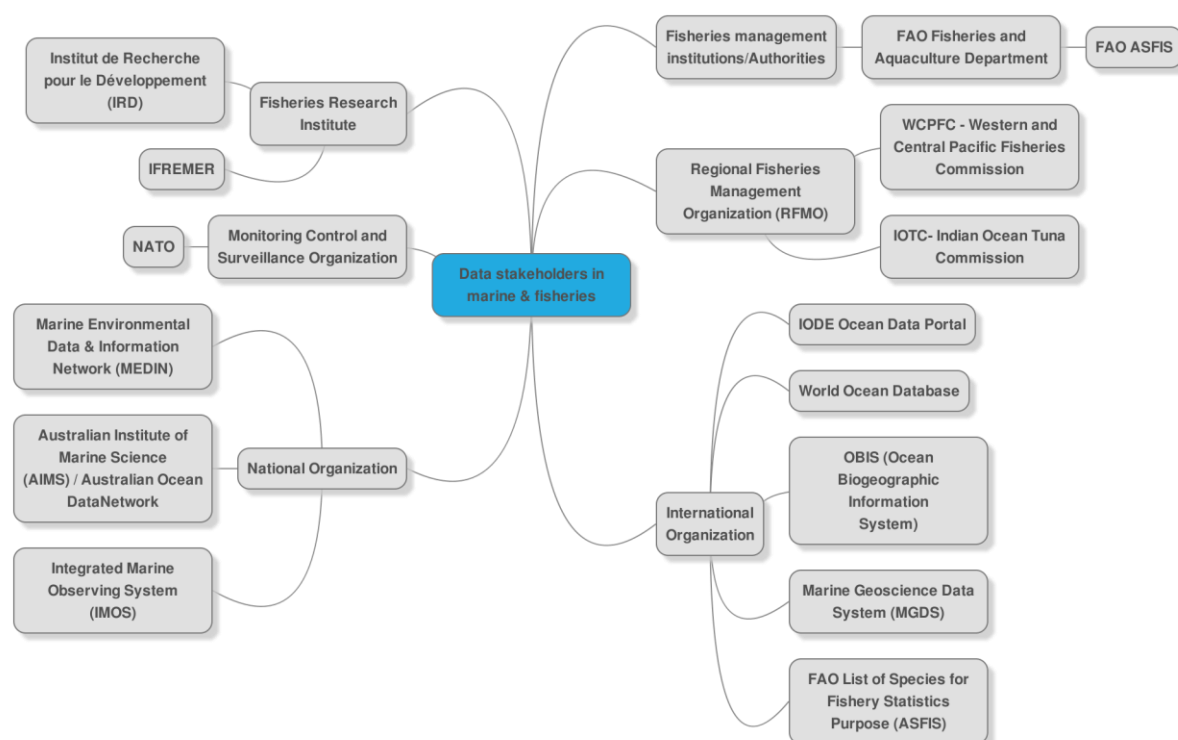


Figure 7: Indicative view of the ecosystem of data stakeholders in marine and fisheries

4.5.1 IODE Ocean Data Portal

The Ocean Data Portal (ODP)⁴⁹ aims at providing seamless access to collections and inventories of marine data from the NODCs (National Oceanographic Data Centres) of the IODE network and allows for the discovery, evaluation (through visualisation and metadata review) and access to data via web services. The data available through the IODE ODP⁵⁰ are mostly real-time hydrometeorological observation data, temperature and salinity data, satellite data about water pollution, aquatic environment etc., as well as EMODNet data. It is one of the three data sources available through IODE.

4.5.2 World Ocean Database

World Ocean Database 2013 (WOD13)⁵¹ is a scientifically quality-controlled database of selected historical in-situ surface and subsurface oceanographic measurements produced by the Ocean Climate Laboratory (OCL) at the National Oceanographic Data Center (NODC), USA. It provides

⁴⁹ <http://www.oceandataportal.org>

⁵⁰ <http://www.oceandataportal.net/portal>

⁵¹ http://www.nodc.noaa.gov/OC5/WOD/pr_wod.html

access to data from two major data sources; Regional Climatologies⁵² and Gulf of Mexico Data Atlas⁵³. It is one of the three data sources available through IODE.

WOD data: WOD Data are available as NetCDF files. Apart from that, they are available as CSV and ASCII, among others.

4.5.3 OBIS (Ocean Biogeographic Information System)

The Ocean Biogeographic information System (OBIS)⁵⁴ seeks to absorb, integrate, and assess isolated datasets into a larger, more comprehensive picture of life in oceans. The system aims to stimulate research about our oceans to generate new hypotheses concerning evolutionary processes, species distributions, and roles of organisms in marine systems on a global scale. The abstract maps that OBIS generates are maps that contribute to the 'big picture' of oceans, by providing a comprehensive, collaborative, world-wide view of oceans.

Over the period 2004–present, an international network of Regional OBIS Nodes has also been established, that are facilitating the connection of data sources in their region to the master OBIS data network and also increasingly provide specialised services or views of OBIS data to users in their particular region.

4.5.4 Marine Geoscience Data System (MGDS)

The Marine Geoscience Data System (MGDS)⁵⁵ provides a suite of tools and services for free public access to marine geoscience research data acquired throughout the global oceans and adjoining continental margins. MGDS provides access to a number of data systems and services, such as the ones mentioned in the following table.

| Name | URL | Scope |
|-----------------------------------|---|---|
| Marine-Geo Digital Library | http://www.marine-geo.org/tools/search/ | Interactive digital data repository and metadata catalogue that offers a suite of services for data discovery and preservation. |
| Community Data Portals | http://www.marine-geo.org/data_portals.php | Data Portals are available for the GeoPRISMS, MARGINS, and Ridge 2000 research initiatives, the Academic Seismic Portal ASP@LDEO and ASP@UTIG collections supporting marine seismic research, the GMRT synthesis of global bathymetry, a collection of submersible data, and a newly initiated Coastal data collection. |
| Global Bathymetry | http://www.marine-geo.org/portals/gmrt | Swath bathymetry data acquired with ships throughout the global oceans are cleaned, quality assured and curated by the our data managers for inclusion in GMRT (Global Multi-Resolution Topography). |
| Antarctic Data Resources | http://www.marine-geo.org/portals/antarctic | MGDS hosts the Antarctic and Southern Ocean Data Synthesis collection of data from research vessels of the US Antarctic Program. It also hosts the US Antarctic Program Data Center (USAP-DC), which provides tools to help scientists find Antarctic data of interest and satisfy their obligation to share data under the NSF Division of Polar Programs (PLR) data policy. |

Table 2: MGDS data sources and services

⁵² http://www.nodc.noaa.gov/OC5/regional_climate

⁵³ <http://gulfatlas.noaa.gov>

⁵⁴ <http://www.iobis.org>

⁵⁵ <http://www.marine-geo.org/index.php>

4.5.5 Marine Environmental Data & Information Network (MEDIN)

The Marine Environmental Data and Information Network (MEDIN)⁵⁶ is a partnership of UK public and private sector organisations committed to improving access to marine data. MEDIN promotes sharing of, and improved access to, these data. Marine data are held by many organisations in the UK and are collected for many different purposes:

- for the timing of tides to determine the position of submerged obstacles
- about the position of submerged obstacles
- for marine conservation
- to monitor and forecast weather and ocean states
- to site marine structures
- for scientific research to understand marine processes

MEDIN aims to provide:

- secure long-term management of marine data sets by setting up a network of Data Archive Centres (DACs)
- improved access to authoritative marine data held in this network, through a central discovery metadata portal
- an agreed set of common standards for metadata, data format and content maintained and supported by partners
- guidelines, contractual clauses and software tools to support standards and best practice data management

4.5.6 Australian Institute of Marine Science (AIMS)

The Australian Institute of Marine Science (AIMS) is Australia's premiere tropical marine research agency. AIMS focuses its work on the following:

- conducting strategic and applied research into marine life, from microbes to whole-of-ecosystems, and the processes that sustain them;
- monitoring the condition and trends in health of the marine environment;
- building models and decision-support tools to assist interpretation of the data collected;
- developing a broad spectrum of enabling technologies, from molecular sciences to ocean technologies.

AIMS has established collaborations with various global marine-related research initiatives, (including NOAA) and plays a key role in national marine policy development, major national and international collaborative frameworks, research and development collaborations, and memorandums of understanding.

In addition, data collected by AIMS and other organisations can be found at the Australian Ocean Data Network⁵⁷ an online network of marine and coastal data resources.

⁵⁶ <http://www.oceannet.org>

⁵⁷ <http://portal.aodn.org.au/aodn>

4.5.7 Integrated Marine Observing System (IMOS)

IMOS⁵⁸ is a national collaborative research infrastructure supported by the Australian Government, led by the University of Tasmania and supported with the Australian marine and climate science community. In order to make specialised data more usable to a wider audience, the IMOS data team and Facilities have developed a number of data tools over the years⁵⁹.

The IMOS Portal web application is an instance of the AODN Open Geospatial Portal⁶⁰. IMOS also hosts IMOS Ocean Current⁶¹, which provides access to data related to ocean information around Australia, such as surface current and temperature.

4.5.8 FAO List of Species for Fishery Statistics Purpose (ASFIS)

The FAO Fisheries and Aquaculture Statistics and Information Service (FIPS) collates world capture and aquaculture production statistics at either the species, genus, family or higher taxonomic levels in 2,189 statistical categories (as of 2013) referred to as species items.

ASFIS list of species includes 12,600 species items selected according to their interest or relation to fisheries and aquaculture. For each species item stored in a record, codes (ISSCAAP group, taxonomic and 3-alpha) and taxonomic information (scientific name, author(s), family, and higher taxonomic classification) are provided. An English name is available for most of the records, and about one third of them have also a French and Spanish name. Information is also provided about the availability of fishery production statistics on the species item in the FAO databases.

4.5.9 FAO Fisheries and Aquaculture Department

The Department promotes policies and strategies aiming at sustainable and responsible development of fisheries and aquaculture in inland and marine waters. For this purpose, the Department provides discussion fora, information, legal and policy frameworks, codes and guidelines, options for strategies, scientific advice, training material, etc. For the purpose:

- It collects, analyses and disseminates information on the sector operations (catch, production, value, prices, fleets, farming systems, employment).
- It develops methodology, assesses and monitors the state of wild resources and elaborates resources management advice.
- It monitors and advises on the development and management of aquaculture.
- It provides socio-economic analysis of fisheries and aquaculture and assists in the elaboration of development and management policies and strategies and institutions.
- It supports and assists a network of regional fishery commissions and promotes aquaculture networks.
- It monitors and advises on technology development, fish processing, food safety and trade.

⁵⁸ <http://imos.org.au>

⁵⁹ <http://imos.org.au/imosdatatools.html>

⁶⁰ <https://github.com/aodn/aodn-portal>

⁶¹ <http://oceancurrent.aodn.org.au>

4.5.10 General Fisheries Commission for the Mediterranean (GFCM)

The General Fisheries Commission for the Mediterranean (GFCM) is a regional fisheries management organisation (RFMO) established under the provisions of Article XIV of the FAO Constitution. The main objective of the GFCM is to promote the development, conservation, rational management and best utilisation of living marine resources as well as the sustainable development of aquaculture in the Mediterranean, the Black Sea and connecting waters. The GFCM is currently composed of 24 members (23 member countries and the European Union) who contribute to its autonomous budget to finance its functioning. Membership is open to Mediterranean coastal States and regional economic organisations as well as to United Nations member States whose vessels engage in fishing in its area of application.

GFCM has recently amended its legal framework and the Agreement⁶² for its establishment with a view to enhancing its efficiency and thus better responding to current and future challenges in the whole region. Among the functions of GFCM is *“to ensure, if possible through electronic means, the collection, submission, verification, storing and dissemination of data and information, consistent with relevant data confidentiality policies and requirements”* – a statement that highlights the importance of data for the Commission.

4.5.11 Joint Research Centre (JRC)

Joint Research Centre⁶³ is EC’s science service. Its mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle. Since 2000, an EU framework for the collection and management of fisheries data is in place. This framework was reformed last in 2008 resulting in the Data Collection Framework (DCF). Under this framework the Member States (MS) collect, manage and make available a wide range of fisheries data needed for scientific advice.

The data is collected on the basis of National Programmes in which the MS indicate which data is collected, the resources they allocate for the collection and how data is collected. MS must report annually on the implementation of their National Programmes and the Scientific, Technical and Economic Committee for Fisheries (STECF) evaluates these Annual Reports. Part of the data collected by the MS is uploaded in databases managed by the JRC in response to data calls issued by DG MARE. This data is analysed by experts of the STECF and forms the basis for scientific opinions and recommendations formulated in STECF reports. The resulting scientific advice is used to inform the CFP decision making process.

JRC is assembling the data, storing it in databases, analysing its quality and coverage and making it available to the STECF working groups. Once the STECF reports are finalised the data is disseminated in aggregated form for a target audience of experts for further use in scientific analyses and policy.

⁶² <http://www.fao.org/gfcm/background/legal-framework/en/>

⁶³ <https://ec.europa.eu/jrc/en>

4.5.12 Global Biodiversity Information Facility (GBIF)

The Global Biodiversity Information Facility (GBIF)⁶⁴ is an international open data infrastructure, funded by governments. It allows anyone, anywhere to access data about all types of life on Earth, shared across national boundaries via the Internet. GBIF operates through a network of nodes, coordinating the biodiversity information facilities of Participant countries and organisations, collaborating with each other and the Secretariat to share skills, experiences and technical capacity.

4.6 Content Analysis of the Fisheries and Marine Data Sources

This section provides an analysis of the data sources described in the previous section. The analysis aims to provide insights on the use of existing standards for the description of data with metadata and their classification with ontologies and other knowledge organisation schemes, such as controlled vocabularies, controlled lists and thesauri, among others.

The use of existing standards for the description of datasets (metadata) as well as for the formulation of the datasets themselves is of highest importance for the interoperability between different data sources and systems. In this context, within the framework of the WMO-IOC Joint Technical Commission for Oceanography and Marine Meteorology (JCOMM), IODE has been cooperating with WMO on the agreement on standards (OceanDataStandards project⁶⁵). The scope of this project was expanded by IODE in 2013 when it established the "Ocean Data Standards and best Practices" project which would also disseminate and promote "best practices", in addition to "standards".

While the standards are published in the IOC Manuals and Guides series, the best practices (and standards) are made available through the new OceanPractices repository managed under the "IODC Clearing House Service for Data/Information Management practices project"⁶⁶.

The following sections provide an overview of the attributes of the data and metadata currently available through some of the most prominent data sources in the context of fisheries, marine and maritime, managed or owned by the key data stakeholders described in the beginning of this chapter.

4.6.1 National Climatic Data Center

NCDC provides access to a wealth of different datasets, as presented in the following table.

| Data source / dataset name | URL | Scope |
|---|---|---|
| Global Marine Data | http://www7.ncdc.noaa.gov/CDO/CDOMarineSelect.jsp | Historical and current marine observations from various national and international sources. |
| International Comprehensive Ocean-Atmosphere Data Set (ICOADS) | http://icoads.noaa.gov/index.shtml | Surface marine data spanning the past three centuries |

⁶⁴ <http://www.gbif.org/>

⁶⁵ <http://www.oceandatastandards.org>

⁶⁶ http://www.iode.org/index.php?option=com_content&view=article&id=374&Itemid=100090

| Data source / dataset name | URL | Scope |
|---|---|---|
| Marine Data Map Access | http://gis.ncdc.noaa.gov/map/viewer/#app=cdo&cfg=marine&theme=marine | Various marine datasets visualised on maps. |
| Merged Land Ocean Surface Temperature Analysis | https://www.ncdc.noaa.gov/data-access/marineocean-data/mlost | Global surface temperature dataset, with monthly resolution from January 1880 to present. |
| NOAA Global Surface Temperature (NOAAGlobalTemp) | https://www.ncdc.noaa.gov/data-access/marineocean-data/noaa-global-surface-temperature-noaaglobaltemp | Combined global sea surface (water) temperature (SST) dataset with global land surface air temperature dataset |
| Surface Flux Analysis (SURFA) | https://www.ncdc.noaa.gov/data-access/marineocean-data/surface-flux-analysis | Model simulations (NWP and climate model output) and observational data (from in situ and satellite platforms). |
| Voluntary Observing Ship Climate (VOSCLIM) Fleet | https://www.ncdc.noaa.gov/data-access/marineocean-data/vosclim | subset of VOS data in both real time and in delayed mode, supplemented by an extensive array of metadata |
| Blended Sea Winds | https://www.ncdc.noaa.gov/data-access/marineocean-data/blended-global/blended-sea-winds | globally gridded, high-resolution ocean surface vector winds and wind stresses |
| Extended Reconstructed Sea Surface Temperature (ERSST) | https://www.ncdc.noaa.gov/data-access/marineocean-data/extended-reconstructed-sea-surface-temperature-ersst | NCEI operational sea surface temperature (SST) analysis product that uses in situ ship and buoy reports to reconstruct historical SSTs from the mid-19th century to present |
| Daily ¼° Optimum Interpolation Sea Surface Temperature (OISST) | https://www.ncdc.noaa.gov/oisst | NCEI operational high-resolution sea surface temperature (SST) analysis product that blends in situ (ship and buoy) reports and satellite data and uses optimum interpolation to fill spatial gaps, produced daily on ¼° grids. |

Table 3: NCDC datasets

In general, data sources provided through NOAA centers use existing metadata standards, as described in the “NAO 212-15: Management of Environmental Data and Information”⁶⁷. Such standards include (but are not limited to) Directory Interchange Format (DIF), Ecological Metadata Language (EML), Sensor Model Language (SensorML), Climate Science Modeling Language (CSML), and NetCDF Markup Language (NcML). In addition, a commonly used metadata standard, the Federal Geographic Data Committee’s (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) was developed in support of the coordinated development, use, sharing, and dissemination of geospatial data on a national basis.

On top of that, NOAA's Environmental Data Management Committee's (EDMC) Data Documentation Planning Directive "establishes ISO 19115 Parts 1 and 2 and a recommended representation standard (ISO 19139) for documenting NOAA's environmental data and information"⁶⁸.

⁶⁷ http://www.corporateservices.noaa.gov/ames/administrative_orders/chapter_212/212-15.html

⁶⁸ <http://www.ncddc.noaa.gov/metadata-standards>

4.6.2 SeaDataNet

SeaDataNet has been working on the establishment of standards for the marine and oceanographic data. Examples are the SeaDataNet NetCDF (CF) data transport model for Marine and Oceanographic Datasets⁶⁹, the SeaDataNet Controlled Vocabularies for describing Marine and Oceanographic Datasets⁷⁰ and the SeaDataNet Common Data Index (CDI) metadata model for Marine and Oceanographic Datasets⁷¹. An overview of the SeaDataNet work on establishing common standards can be found on the Ocean Data Standards portal⁷².

SeaDataNet Metadata: As regards metadata, all SeaDataNet metadata services (CSR, CDI, EDMED, EDMERP and EDIOS) make use of XML formats and exchange schema's (XSD), based on the ISO 19115 content model. On top of that, the SeaDataNet data are described using common vocabularies based on the NERC DataGrid (NDG) Vocabulary Server.

SeaDataNet Data: As regards data, data sets are accessible via SeaDataNet download services in the following data transport formats:

- SeaDataNet ODV4 ASCII for profiles, time series and trajectories,
- SeaDataNet NetCDF with CF compliance for profiles, time series and trajectories,
- SeaDataNet MedAtlas as optional extra format,
- NetCDF with CF compliance for 3D observation data such as ADCP.

The SeaDataNet NetCDF (CF) format for profiles, time series and trajectories has been defined by bringing together a community comprising NetCDF and CF experts (such as from NCAR and UNIDATA), and as many users of CF NetCF for oceanographic point data as possible.

4.6.3 OBIS

The Ocean Biogeographic Information System⁷³ provides access to marine species datasets from all of the world's oceans. The OBIS database repository allows users to identify biodiversity hotspots and large-scale ecological patterns, analyse dispersions of species over time and space, and plot species' locations with temperature, salinity, and depth. The OBIS Portal also hosts, or provides links to, a variety of software tools that can operate on OBIS data to provide mapping, analysis, or data modelling services, including the KGS Mapper, the c-squares mapper, and the ACON mapper.

4.6.4 MGDS

Different data products of MGDS⁷⁴ use different standards; each Data Portal includes an interactive map to facilitate data discovery as well as a customised search options with data portal-specific search options and results. For example, the Marine-Geo Digital Library which

⁶⁹

http://www.oceandatastandards.org/images/stories/ods_docs/proposal%20201502_SDN2_D89_WP8_ODSBP_SeaDataNet_NetCDFCF_Proposal.pdf

⁷⁰ <http://www.seadatanet.org/Standards-Software/Common-Vocabularies>

⁷¹ http://www.oceandatastandards.org/images/stories/ods_docs/prop201401_ODSBP-SeaDataNet%20CDI%20Data%20Model%20proposal.doc

⁷² <http://www.oceandatastandards.org/submitted-proposals-mainmenu-49>

⁷³ <http://iobis.org/mapper>

⁷⁴ http://www.marine-geo.org/data_portals.php

provides open web-based access to 56.8TB of data, corresponding to over 694,000 digital data files from more than 2,695 research programs,

4.6.5 MEDIN

The Data Discovery Portal of MEDIN⁷⁵ is a metadata discovery service providing users with a single point of access to individual metadata records submitted to the portal by the Data Archive Centres and other public and private sector bodies. The portal provides access to marine datasets and commonly required marine reference layers. Metadata records are available for UK marine data sets across all subject areas and disciplines. This includes UK organisations undertaking data collection in non-UK waters. There are other portals which contain data for more specific data themes or from a European rather than a UK perspective.

The main sources of metadata are the network of MEDIN accredited Data Archiving Centres (DACs). These currently include geophysical and geological data from the British Geological Survey (BGS), oceanographic data from the British Oceanographic Data Centre (BODC), bathymetry data from the UK Hydrographic Office (UKHO) and habitat and species data from the Data Archive on Seabed Species and Habitats (DASSH).

MEDIN has worked on the establishment and adoption of a number of standards for the description and interoperability of marine data. These standards include:

- MEDIN Discovery Metadata Standard⁷⁶, a marine profile of the UK government Standard GEMINI2 which is also compliant with other international conventions such as INSPIRE and ISO19115.
- MEDIN controlled vocabularies⁷⁷. The following vocabularies are used by the MEDIN portal.

| Controlled Vocabulary | Details |
|---|---|
| Chemical, biological or physical parameter | <u>BODC Parameter Usage Vocabulary</u> |
| Development stage | <u>ICES Reference Codes</u> |
| Keywords | <u>SeaDataNet Parameter Discovery Vocabulary</u> |
| UK Species | <u>The Marine Species of the British Isles and Adjacent Seas (MSBIAS)</u> |
| World Species | <u>World register of Marine Species (WoRMS)</u> |

Table 4: Controlled vocabularies used by the MEDINE

- Other standards used: MEDINE also considers the adoption of existing standards for other applications, including the following:
 - Marine Survey Data Management Handbook v1. British Geological Survey
 - BGS Survey Folder Structure
 - Mapping European Seabed Habitats (MESH) Guidelines
 - IHO standard for Hydrographic Surveys, Special Publication S-44
 - ICES multibeam echosounder data

⁷⁵ <http://portal.oceannet.org/search/full>

⁷⁶ http://www.oceannet.org/marine_data_standards/medin_disc_stdn.html

⁷⁷ http://www.oceannet.org/marine_data_standards/controlled_vocababs.html

4.6.6 ICES

The ICES dataset collections are described with metadata based on the ISO 19139 specification which uses advanced XML techniques to add capabilities to ISO metadata, related to identifying and referring to "objects" in the metadata. The ICES Data Centre accepts a wide variety of marine data and meta-data types into its databases. The data formats, guidelines and vocabularies are specific to the type of data and whether it is associated with a marine convention monitoring programme.

The ICES Vocabulary Server⁷⁸ is the reference codes library for trawl survey, oceanographic, commercial fishery and environmental data. It contains both ICES specific codes as well as external codes.

4.6.7 AIMS Data

The data that AIMS collects⁷⁹ are described in detail in the AIMS Data Catalogue. The catalogue contains descriptions of data (often called metadata), which allows others, including researchers, to identify potential data of use to them. Data can then be downloaded (if available) or requested. Access to some data will be via legal data agreement.

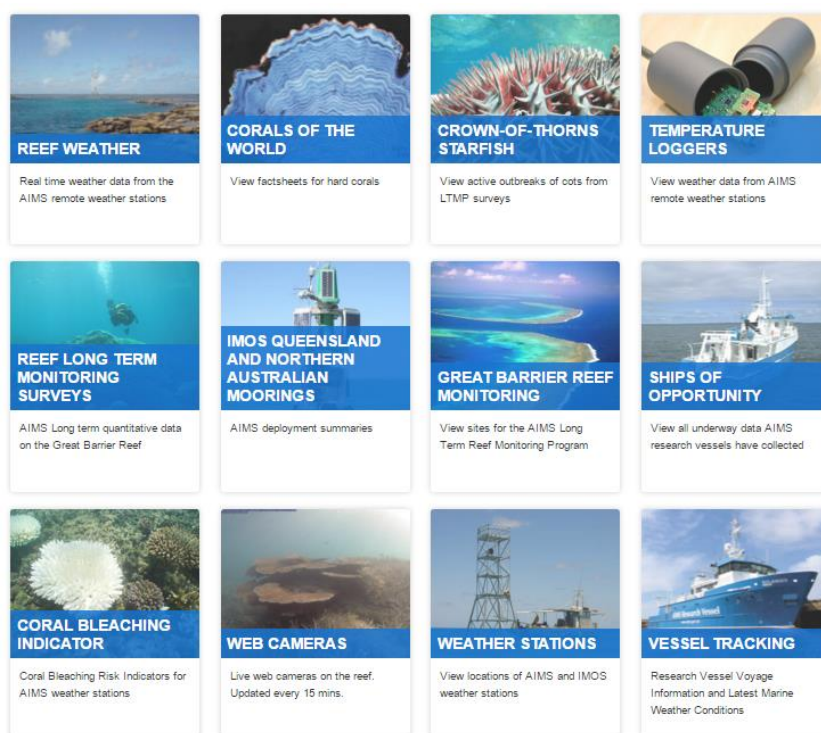


Figure 8: The AIMS data portal

Data collected by AIMS and other organisations can be found at the Australian Ocean Data Network, an online network of marine and coastal data resource.

⁷⁸ <http://vocab.ices.dk>

⁷⁹ <http://www.aims.gov.au/docs/data/data.html>

4.6.8 AODN Ocean Data Portal

The AODN Ocean Data Portal⁸⁰ is a single access point for marine data published by Australian Commonwealth Agencies and by a large number of other data contributors. It provides access to the complete IMOS metadata catalogue (described in the previous section) and all available ocean data⁸¹. The AODN Portal includes data from the six Commonwealth Agencies with responsibilities in the Australian marine jurisdiction.

AODN data, available as map layers, are organised in four (4) categories: Observation Data (from various sources), Model Outputs, Data Products and Satellite Data. AODN datasets are available as NetCDF files.

4.6.9 IMOS

All IMOS data can be discovered, accessed and downloaded via the IMOS Ocean Portal⁸². There is a wide range of marine data types available in a variety of formats, organised by Parameter (Physical/Water, Biological, Physical/Atmosphere and Chemical), Organisation, Type of Platform, Date, Geographic Boundary and Keyword. Each dataset is described with metadata using existing standards, such as the Marine Community Profile (MCP) v2.0, which is a subset of the international standard and includes all ISO 19115 core metadata elements, as well as supplementary elements, codelists and vocabularies to assist in the description of marine resources. Three subsets of MCP are supported:

1. MCP Minimum: Contains the Data Identification and Metadata Info sections (groups) of the Marine Community Profile.
2. MCP Core: The core metadata components of ISO19115 and,
3. MCP All: An expanded version of the MCP v1.4 that is also compliant with ISO 19139.

Metadata records are created using a GeoNetwork Metadata Entry and Search Tool (MEST).

IMOS non-gridded data is served by OGC WFS servers. Gridded data is served by the GoGoDuck.

4.6.10 FAO ASFIS

Data available through ASFIS is classified according to the FAO 'International Standard Statistical Classification for Aquatic Animals and Plants' (ISSCAAP) which divides commercial species into 50 groups on the basis of their taxonomic, ecological and economic characteristics. This classification is further refined through the use of taxonomic codes and 3-alpha identifiers.

ASFIS data can be downloaded from the FAO website as a compressed (.zip) file; it is a .txt file in delimited format, which can be imported to most existing spreadsheet or database software. ASFIS data can also be searched and retrieved online through the FAO TERM Portal⁸³ developed by the FAO Terminology Team.

⁸⁰ <http://portal.aodn.org.au/aodn>

⁸¹ https://imos.aodn.org.au/data_collections.html

⁸² <https://imos.aodn.org.au/imos123>

⁸³ <http://www.fao.org/faoterm/en/>

4.6.11 FAO Fisheries and Aquaculture Department

The Fisheries and Aquaculture Department of UN FAO provides access to a variety of data types, such as statistics, geographical information, publications and fact sheets. As regards the statistical data, they can be accessed through (i) FishStatJ - Software for fishery statistical time series, (ii) Online Query Panels and (iii) online YearBooks.

A full list of the statistical collections can be found online⁸⁴. These datasets are accompanied with basic information such as collection overview, available formats and information products and status of the collection.

The list of Fisheries and Aquaculture Fact Sheets include geographic profiles, technology documentation, information standards as well as other resources. These factsheets are available as web pages (HTML/XML). Geographical information consists of various types of maps as well as geographic profiles. Last but not least, publications include technical papers, proceedings, reports, guidelines and year books, among others, all available in PDF format.

Publications are described with metadata, which provide basic information about the document (e.g. Title, Abstract, Publication year and type).

4.6.12 FAO GFCM

Through the GFCM portal⁸⁵, a number of different datasets can be accessed providing information on various aspects related to fisheries, such as fleets, ports, maps and fish stock, as well as experts and institutions activated in the fisheries' context. Data are available in the form of spreadsheets that are automatically visualised online in the form of charts and diagrams. In various cases, a data browser is also available/embedded, allowing users to access primary data.

GFCM data can also be accessed through other locations, such as the Fisheries and Aquaculture Department of FAO which provides access to the Capture Production in the Mediterranean and Black Sea using GFCM data⁸⁶.

4.6.13 Joint Research Centre (JRC) fisheries data collection

The Joint Research Centre (JRC) fisheries data collection web site⁸⁷ provides access to various types of data and information on EU aquaculture and fisheries, including:

- latest news in relation to data calls, deadlines, variable definitions, disaggregation levels and uploading procedures;
- National Programs and Annual Report prepared by the MS;
- access to the uploading facilities and data dissemination platforms for the experts and the general public;
- coverage reports on the data provided by the MS in response of the data calls managed by JRC;
- DCF technical documents, guidelines and legislation.

⁸⁴ <http://www.fao.org/fishery/statistics/collections/en>

⁸⁵ <http://www.fao.org/gfcm/data/capture-production-statistics/en/>

⁸⁶ <http://www.fao.org/fishery/statistics/gfcm-capture-production/en>

⁸⁷ <https://datacollection.jrc.ec.europa.eu/>

What makes this data source stand out from the rest described in this document is its user-centric approach; registered users are allowed to upload data in the form of spreadsheets, based on the data transmission template provided by JRC.

4.6.14 GBIF

The data accessible through GBIF relate to evidence about more than 1.6 million species, collected over three centuries of natural history exploration and including current observations from citizen scientists, researchers and automated monitoring programmes. As of March 2016, GBIF provides access to 15,640 datasets from 800 data publishers, referring to more than 1,6 million species and 647 million occurrences and 570 million georeferenced records.

The data available through GBIF are metadata records according to the Darwin Core (DwC) metadata standard and the GBIF metadata profile and occurrence data. GBIF data are available through the GBIF RESTful JSON based API⁸⁸ which provides registration, discovery and access and information services.

In addition to the RESTful JSON API, Datasets are exposed using OAI-PMH⁸⁹. Two metadata formats can be retrieved: Ecological Metadata Language (EML) and OAI Dublin Core. Datasets are grouped into sets according to type, country and installation.

Detailed documentation on the GBIF data interoperability options can be found online⁹⁰.

4.7 Data flows

In order to illustrate the legal interoperability framework, several generic data flows were identified where legal interoperability is relevant. The examples slowly zoom out of the individual data to a more service level perspective. Each of these flow perspectives has overlapping legal interoperability characteristics, and together these offer a comprehensive view on the issues.

The dataflows were used in the stakeholder consultation, and experts were asked to respond to more or less theoretical examples legal interoperability issues surrounding data exchange.

Data interoperability issues often overlap with legal data interoperability. This Deliverable limits the perspective to those situations where a legal instrument is needed or envisionable to present a framework solution supported by infrastructure services. Many stakeholder issues were excluded from this Deliverable, as they are beyond the EGI Engage D2.6 scope; issues such as outdated formats or inaccessible repositories, different aggregation levels and procedures, poor quality of (meta)data, incompatible usage policies and separate infrastructures are all known to impede data use.

4.7.1 Data usage and citation

The legal interoperability issues related to direct exploitation of data:

⁸⁸ <http://api.gbif.org/v1/>

⁸⁹ <http://api.gbif.org/v1/oai-pmh/registry>

⁹⁰ <http://www.gbif.org/developer/registry>

- **Data usage;** most practitioners use data to e.g. populate stock population assessment models. They need to be able to locate and access the data, and thus need policies that inform them on the data availability.
- **Citation;** practitioners using data need to be able to attribute the data of others, and clear instructions and tools to properly cite data and owners are valuable.
- **Data formats;** the transformation of data from one format to another, even within the same domain, is an often cumbersome and sometimes impossible task in the absence of a legal framework that enforces a comprehensive description of the format used, including the provenance and quality of the data contained.

This segment of data flows has traditionally been well covered from a legal perspective, but in reality often it is not really enforced. The rapid advance of monitoring and accounting technologies now raises the opportunity to have much finer grained tracking and enforcement options, but it also presents an opportunity to automate a large part of relevant legal interoperability instructions, such as generation of quality metadata, management of citation and references, and timely release of datasets.

4.7.2 Data management

At the level of data management, the proper application of the existing legal instruments to enhance legal interoperability scenarios. Data management here implies that the owner of the data defines the use of data with regards to:

- **Terms of Use;** the rules that users must agree to abide by in order to use a service, define how the data manager wants to manage the use of the information resource. The terms can be merely a disclaimer, especially regarding the use of websites, or cover a more comprehensive set of data exchange conditions, including legal interoperability issues.
- **Copyright;** the selection of the proper copyright to attach to a dataset or publication is receiving much attention. Most data managers implement the copyright policy of their organisations, but are increasingly aware of, or even confronted with, the necessity to adapt to modern data requirements for service level ownership, sharing agreements, derivative products, etc. Reconciling different copyrights at service level is a known concern.
- **Data Policy;** The data policies cover the terms of use and copyright issues, but also include best practices, data preparation and curation, sustainability and governance. Here, the questions are beyond the use and management of data, but address the strategic interest of the organisations and data owners in the long term.

The development of a comprehensive legal framework for data management is currently the focus of much debate, e.g. in the RDA Legal interoperability WG.

4.7.3 Infrastructure support to data flows

At infrastructure level, when managing data flows, factors that impact on legal interoperability are related to the governance and long term sustainability of the data.

- **Data ownership;** at the level of an infrastructure, ownership issues can bring new requirements and roles to an institution management of their data, and e.g. necessitate the

establishment of a custodian or trusted data manager outside the organisation, which may complicate the legal framework defining ownership.

- **Data location;** Most owners currently refer to a custodian within the (often literally) walls of an organisation. Reformulating an organisation mandate on data can require a lengthy and complicated negotiation, especially if there are multiple data contributors.
- **Data strategy;** most datasets (at least in the fisheries domain) are established to meet specific objectives related to the management of fisheries resources. By design, the strategy to manage these datasets are not service oriented, and the implementation of more Open Data and infrastructure services based strategies thus requires careful attention.

The establishment of EU and global data infrastructures necessitates reconsideration at organisational and collaborative level on how best to leverage the advantages of infrastructure based services to meet new requirements for legal interoperability.

4.8 Legal Analysis of the Fisheries and Marine Data sources

Our analysis so far showed that different data sources apply different types of licensing schemas or use disclaimers over the data that they produce or manage. These may come in the form of CC licenses, Legal Notices, Disclaimers, Copyright texts etc. For facilitating the identification of common patterns in these different types of licensing “statements”, we have worked towards the organisation of the available information in tables that are based on the CC. Despite the fact that effort was made to map as precisely as possible the information available in the various legal statements, this mapping may have to be considered as experimental at this point.

4.8.1 IODE

The UNESCO/IOC Oceanographic Data Exchange Policy⁹¹ indicates, “Member States shall provide timely, free and unrestricted access to all data, associated metadata and products generated under the auspices of IOC programmes”. The same applies to data and metadata “from non-IOC programmes that are essential for application to the preservation of life, beneficial public use and protection of the ocean environment, the forecasting of weather, the operational forecasting of the marine environment, the monitoring and modelling of climate and sustainable development in the marine environment”. In both cases, these apply for non-commercial use by the research and education communities, provided that any products or results of such use shall be published in the open literature without delay or restriction.

| | |
|--|---|
| Copyright and other Restrictions (Y/N) | N |
| Are commercial uses of this resource allowed? (Y/N) | N |
| Are modifications of your work of this resource by other people allowed? (Y/N/Yes, if others share alike) | Y |
| Cost (Y/N) | N |
| Description | Available at www.iode.org/policy |

⁹¹ www.iode.org/policy

Various data sources around the world, including members of the IODE network such as the Hellenic National Oceanographic Data Centre⁹² (but not limited to them) have adopted the IODE Oceanographic Data Exchange Policy.

4.8.2 NOAA / NCEI Centers

The data available through all NOAA/NCEI Centers are covered by the same disclaimer, which is provided in the table below.

| | |
|--|---|
| Copyright and other Restrictions (Y/N) | N |
| Are commercial uses of this resource allowed? (Y/N) | N/A |
| Are modifications of your work of this resource by other people allowed? (Y/N/Yes, if others share alike) | N/A |
| Cost (Y/N) | Depends |
| Description | <p>NCEI makes every effort to place most datasets online where they can be obtained for free. However, we must recover the costs for producing CD-ROM & DVD products and some publications. Also, having a customer service representative do a customised retrieval for you will also have some fee involved. See our Ordering and Payment Procedures. CD-ROMs & DVDs and Publications can also be ordered from the Online Store at a discount.</p> <p>This website is provided as a public service by the U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data and Information Service. Information presented on these web pages is considered public information and may be distributed or copied. Use of appropriate byline/photo/image credit is requested.⁹³</p> |

4.8.3 SeaDataNet

SeaDataNet has defined an overarching SeaDataNet Data policy⁹⁴ that aims to strike a balance between the rights of investigators and the need for widespread access through the free and unrestricted sharing and exchange of SeaDataNet data, meta-data and data products. The final goal of this policy is to serve the scientific community, public organisations, and environmental agencies, and to facilitate the production of advice and status reports by stating the conditions for data submission, access and use. This policy applies to data managed by SeaDataNet partners for providing access to data managed in the SeaDataNet distributed systems.

| | |
|--|---|
| Copyright and other Restrictions (Y/N) | N |
| Are commercial uses of this resource allowed? (Y/N) | ? |
| Are modifications of your work of this resource by other people allowed? (Y/N/Yes, if others share alike) | ? |

⁹² http://hnodc.hcmr.gr/?page_id=66

⁹³ <http://www.nodc.noaa.gov/about/faq.html>

⁹⁴ <http://www.seadatanet.org/content/download/3899/29604/file/SeaDataNet%20Data%20Policy%20.pdf>

| | |
|--------------------|---|
| Cost (Y/N) | N |
| Description | Data Users must respect any and all restrictions on the use or reproduction of data. The use or reproduction of data for commercial purpose might require prior written permission from the data source. Data Users should not give to third parties any SeaDataNet data or product without prior consent from the source Data Centre. |

4.8.4 OBIS

OBIS is one of the data sources that provide detailed guidelines on the sharing and use of data in OBIS⁹⁵. These guidelines build on the IODE Data Exchange Policy but go into depth in terms of Data Use and Data Sharing. As regards the latter, the guidelines refer that “The data providers retain all rights and responsibilities associated with the data they make available to OBIS via the OBIS nodes. The OBIS nodes warrant that they have made the necessary agreements with the original data providers that it can make the data available to OBIS data under the following CC licenses: CC-0 or CC-BY or CC-BY-NC), of which CC-0 is the preferred one and CC-BY-NC the least preferred. The data providers are responsible for the completeness of the data and metadata profiles.”

| | |
|--|--|
| Copyright and other Restrictions (Y/N) | N |
| Are commercial uses of this resource allowed? (Y/N) | Depends on the licensing schema applied |
| Are modifications of your work of this resource by other people allowed? (Y/N/Yes, if others share alike) | Y |
| Cost (Y/N) | N |
| Description | The data providers retain all rights and responsibilities associated with the data they make available to OBIS via the OBIS nodes. The OBIS nodes warrant that they have made the necessary agreements with the original data providers that it can make the data available to OBIS data under the following CC licenses CC-0 or CC-BY or CC-BY-NC), of which CC-0 is the preferred one and CC-BY-NC the least preferred. The data providers are responsible for the completeness of the data and metadata profiles. |

4.8.5 MGDS

The Terms of Use of MGDS⁹⁶ indicate that “Data and metadata are licensed under a CC Attribution-Noncommercial-Share Alike 3.0 United States License. Appropriate acknowledgment with a byline/credit/link must be given to both the original scientists/data contributors by reference to their relevant publications and to the Marine Geoscience Data System (www.marine-geo.org). Where citation information has been provided to us by scientists it is included with the relevant database entries, and should be acknowledged when data are used. You may browse freely, but you may not circulate or publish materials you obtained from this site if you do not accept the terms of providing adequate citation.”

⁹⁵ <http://www.iobis.org/node/639>

⁹⁶ http://www.marine-geo.org/about/terms_of_use.php

| | |
|--|--|
| Copyright and other Restrictions (Y/N) | N |
| Are commercial uses of this resource allowed? (Y/N) | Depends on the licensing schema applied |
| Are modifications of your work of this resource by other people allowed? (Y/N/Yes, if others share alike) | ShareAlike |
| Cost (Y/N) | N |
| Description | Data and metadata is licensed under a CC Attribution-Noncommercial-Share Alike 3.0 United States License. Appropriate acknowledgment with a byline/credit/link must be given to both the original scientists/data contributors by reference to their relevant publications and to the Marine Geoscience Data System (www.marine-geo.org). Where citation information has been provided to us by scientists it is included with the relevant database entries, and should be acknowledged when data are used. You may browse freely, but you may not circulate or publish materials you obtained from this site if you do not accept the terms of providing adequate citation |

4.8.6 MEDIN

Information related to the use, reuse and cost of a dataset available through the MEDIN Data Portal are provided at a dataset level, in the “Conditions for access and use constraints” metadata element. Access to a dataset may be restricted by the dataset provided; in this case, only the metadata are available for the specific dataset.

| | |
|--|--|
| Copyright and other Restrictions (Y/N) | Depends on the specific dataset licensing |
| Are commercial uses of this resource allowed? (Y/N) | Depends on the specific dataset licensing |
| Are modifications of your work of this resource by other people allowed? (Y/N/Yes, if others share alike) | Depends on the specific dataset licensing |
| Cost (Y/N) | Depends on the specific dataset licensing |
| Description | Available at a dataset level through the metadata description of each dataset. |

4.8.7 ICES

ICES is working on facilitating access to and reuse of the data managed by ICES. The ICES Data Policy⁹⁷ indicates that data users can obtain publicly available data as soon as is feasible. Data sources must be acknowledged, preferably using a formal citation (templates for citation are provided in the policy document) and as regards potential applications, data users must respect all restrictions on the use of data such as for commercial purposes. As regards redistribution of data, data may only be redistributed, i.e., made available in other data collections or data portals, with the prior written consent of ICES.

As regards the Access and use conditions for Vessel Monitoring System (VMS) data made available through ICES Data calls, the conditions that apply are described in a different document⁹⁸. The document explicitly mentions that such data shall be used only for the purposes of facilitating scientific advice. The data should be treated as confidential and the transmission or sharing of

⁹⁷ http://ices.dk/marine-data/Documents/ICES_Data_Policy_2012.pdf

⁹⁸ http://ices.dk/marine-data/Documents/VMS_DataAccess_ICES.pdf

these data are not allowed. The document also mentions that data sources must be duly acknowledged when such data is used, and that end users must respect any and all restrictions on the use or reproduction of data such as restrictions on use for commercial purposes.

| | |
|--|---|
| Copyright and other Restrictions (Y/N) | Y |
| Are commercial uses of this resource allowed? (Y/N) | N |
| Are modifications of your work of this resource by other people allowed? (Y/N/Yes, if others share alike) | N |
| Cost (Y/N) | N |
| Description | http://ices.dk/marine-data/Documents/ICES_Data_Policy_2012.pdf |

4.8.8 AIMS

The content available through the AIMS data portal is licensed under a CC Attribution license. Using the Attributing AIMS page⁹⁹ and the Copyright Notice¹⁰⁰ were operational.

| | |
|--|---|
| Copyright and other Restrictions (Y/N) | N |
| Are commercial uses of this resource allowed? (Y/N) | Y |
| Are modifications of your work of this resource by other people allowed? (Y/N/Yes, if others share alike) | N |
| Cost (Y/N) | N |
| Description | http://www.aims.gov.au/docs/cc-attribution.html ©Australian Institute of Marine Science http://www.aims.gov.au/docs/cc-copyright.html |

4.8.9 IMOS & AODN

IMOS makes all of its data freely available to the marine and climate science community, other stakeholders and users, and international collaborators¹⁰¹. In this context, IMOS data is made freely available under the Conditions of Use¹⁰². Both IMOS data and the contents of its site are licensed under a CC Attribution 4.0 International License. The same licensing schema applies to the data available through the AODN portal.

| | |
|--|---|
| Copyright and other Restrictions (Y/N) | N |
| Are commercial uses of this resource allowed? (Y/N) | Y |
| Are modifications of your work of this resource by other people allowed? (Y/N/Yes, if others share alike) | Y |
| Cost (Y/N) | N |
| Description | Any users of IMOS data are required to clearly acknowledge the source of the material derived from IMOS in the format: Acknowledgement: "Data was sourced from the Integrated Marine Observing System (IMOS) - IMOS is a national collaborative research infrastructure, |

⁹⁹ <http://www.aims.gov.au/docs/cc-attribution.html>

¹⁰⁰ <http://www.aims.gov.au/docs/cc-copyright.html>

¹⁰¹ <http://imos.org.au/imostermsfuse0.html>

¹⁰²

http://www.imos.org.au/fileadmin/user_upload/shared/IMOS%20General/documents/internal/IMOS_Policy_documents/Policy-Acknowledgement_of_use_of_IMOS_data_26Mar15.pdf

supported by Australian Government.”

4.8.10 FAO ASFIS

No information related to the use, reuse and sharing of ASFIS data specifically is provided through the ASFIS website. A bibliographic citation is provided automatically to those interested in referencing ASFIS as the source of the data. However, through the website there is ©FAO link, which leads to the main Terms and Conditions¹⁰³ web page of FAO, where detailed information is provided on the use, reuse and modification terms for the content available through the FAO services.

| | |
|--|---|
| Copyright and other Restrictions (Y/N) | Y |
| Are commercial uses of this resource allowed? (Y/N) | N |
| Are modifications of your work of this resource by other people allowed? (Y/N/Yes, if others share alike) | Y, with appropriate citation: <i>This is an adaptation of an original work by FAO. Views and opinions expressed in the adaptation are the sole responsibility of the author or authors of the adaptation and are not endorsed by FAO.</i> |
| Cost (Y/N) | N |
| Description | <p>Content on the FAO website is protected by copyright. To ensure wide dissemination of its information, FAO is committed to making its content freely available and encourages the use, reproduction and dissemination of the text, multimedia and data presented. Content may be adapted, translated, copied, printed and downloaded for private study, research and teaching purposes, and for use in commercial and non-commercial products or services, provided that appropriate acknowledgement of FAO as the source is given and that FAO's endorsement of users' views, products or services is not stated or implied in any way. FAO encourages unrestricted use of news releases provided on the FAO website, and no formal permission is required to reproduce these materials.</p> <p>RECOMMENDED CITATION: [© FAO] [Year of publication] [Title of content] [Page number (for publications)] [Location on FAO website] [Date accessed and/or downloaded]</p> |

4.8.11 FAO Fisheries and Aquaculture Department

There is no specific information regarding the access, use and sharing policy of the GFCM data; as a result, the default FAO Terms of Use (as described in the previous section) apply in the case as well.

4.8.12 FAO GFCM

There is no specific information regarding the access, use and sharing policy of the GFCM data; as a result, the default FAO Terms of Use (as previously described) apply in the case as well.

¹⁰³ <http://www.fao.org/contact-us/terms/en/>

4.8.13 JRC

Data available through JRC are available to end users so that they can be reused for (i) fisheries management, (ii) policy development and (iii) scientific publication. Such data are publicly available through EU member states based on the Article 18 of Council Regulation (EC) No 199/2008 (25/2/2008). The use of aggregated data already published in STEFC reports is possible without asking for authorisation. No other information on the access, use and sharing of data available through JRC is available through the JRC portal.

| | |
|--|---|
| Copyright and other Restrictions (Y/N) | N |
| Are commercial uses of this resource allowed? (Y/N) | N |
| Are modifications of your work of this resource by other people allowed? (Y/N/Yes, if others share alike) | ? |
| Cost (Y/N) | N |
| Description | |

4.8.14 iMarine

Despite the fact that iMarine cannot be considered as a data provider, the project has worked on a document titled “iMarine Policy and Guidelines” describing its data access and sharing policy.

As regards the copyright and licensing of the content available through the iMarine e-infrastructure, all content on the Infrastructure is subject to copyright. Reuse and licensing terms are defined at a metadata record level, contained in the copyright notices and business metadata accompanying each item of content or dataset (see “licence” or “constraints” field).

Except where otherwise indicated, content is licensed for display and reuse via the Infrastructure under the CC (CC) licence. All derivative products intended for publication and dissemination through iMarine shall be licensed under a CC licence. In addition to providing the licensing metadata required by the CC licence, users must ensure accurate attribution and provide a recommended citation.

| | |
|--|---|
| Copyright and other Restrictions (Y/N) | Y |
| Are commercial uses of this resource allowed? (Y/N) | Depending on the licensing of each content item |
| Are modifications of your work of this resource by other people allowed? (Y/N/Yes, if others share alike) | Depending on the licensing of each content item |
| Cost (Y/N) | N |
| Description | |

4.8.15 Eurostat

Eurostat has a policy¹⁰⁴ of encouraging free re-use of its data, both for non-commercial and commercial purposes. All statistical data, metadata, content of web pages or other dissemination tools, official publications and other documents published on its website, with the exceptions listed below, can be reused without any payment or written licence provided that:

- the source is indicated as Eurostat;
- when re-use involves modifications to the data or text, this must be stated clearly to the end user of the information.

¹⁰⁴ <http://ec.europa.eu/eurostat/about/our-partners/copyright>

The permission granted above does not extend to any material whose copyright is identified as belonging to a third-party, such as photos or illustrations from copyright holders other than the European Union. In these circumstances, authorisation must be obtained from the relevant copyright holder(s). On top of that, there are Eurostat datasets and documents may not be reused for commercial purposes (but non-commercial reuse is possible without restriction).

| | |
|--|---|
| Copyright and other Restrictions (Y/N) | Y (depending on dataset) |
| Are commercial uses of this resource allowed? (Y/N) | Y (with the exclusion of specific datasets) |
| Are modifications of your work of this resource by other people allowed? (Y/N/Yes, if others share alike) | Y, as long as the proper statement is made |
| Cost (Y/N) | N |
| Description | |

4.8.16 GBIF

Following the Secretariat's community consultation in 2013 and 2014, the Governing Board established a general policy to ensure that all species occurrence datasets within the network are associated with digital licenses equivalent to one of the following three choices supplied by CC:

- CC0, under which data are made available for any use without restriction or particular requirements on the part of users
- CC-BY, under which data are made available for any use provided that attribution is appropriately given for the sources of data used, in the manner specified by the owner
- CC-BY-NC, under which data are made available for any use provided that attribution is appropriately given and provided the use is not for commercial purposes

GBIF strongly encourages data publishers to opt for the most open choice available (CC0).

Biodiversity data accessible via the GBIF network are openly and universally available to all users within the framework of the GBIF Data Use Agreement and with the terms and conditions that the Data Publisher has identified in its metadata. Data retrieved from the GBIF network shall be cited according to the "dataset citation provided by the publisher", as shown on the dataset or occurrence page on the GBIF portal. If the publisher-provided citation is either missing or incomplete, the user shall observe the "default citation" given on the dataset or occurrence page.

| | |
|--|---|
| Copyright and other Restrictions (Y/N) | N |
| Are commercial uses of this resource allowed? (Y/N) | Depends on dataset |
| Are modifications of your work of this resource by other people allowed? (Y/N/Yes, if others share alike) | Y |
| Cost (Y/N) | N |
| Description | http://www.gbif.org/terms/licences/ |

4.9 The bigger picture

The European Commission is particularly interested in the exploitation of marine data from various sources for facilitating research as well as for helping various types of stakeholders (such as industry, public authorities and researchers) find and access data so that they can be used for the development of data-powered products and services. In the context of Marine Knowledge

2020¹⁰⁵, the integration of different national and local systems is expected to significantly contribute to this aim through prototypes like the European Marine Observation and Data Network (EMODNET)¹⁰⁶ data portals, a marine data initiative from the European Commission Directorate-General for Maritime Affairs and Fisheries (DG MARE). Due to the fact that the availability of the data used in the EMODNET portals needed to be ensured in terms of access, use and reuse, the issue of legal interoperability was one of the main issues studied by the network.

As regards the EU fisheries management, as it also heavily relies on data collected, managed and supplied by EU countries under the Data Collection Framework. A document titled *“Towards a new Union Framework for collection, management and use of data in the fisheries sector and support for scientific advice regarding the Common Fisheries Policy”*¹⁰⁷ (currently available as a working document) highlights the role of data in the fisheries’ context for research and other purposes (such as policy making); however, the legal interoperability between the various data sources mentioned throughout the document is not discussed. On the other hand, DG MARE’s *“Evaluation of the Data Collection Framework (DCF)”*¹⁰⁸ mentions licensing as one of the barriers in the fisheries’ sector; a fact that highlights the need for enhancing legal interoperability in the fishery and marine context.

4.10 Stakeholder Findings Representing Opportunities for EGI

The above stakeholder classification and problem statements were used in the analysis of the needs for infrastructure services in support of legal interoperability in the (mainly) Fishery and (some) Marine Sciences Data. The Stakeholder consultation evidenced that Legal Interoperability in the fisheries sector or stakeholders relates mainly to practical data management issues and restrictive mandates, whereas from a future infrastructure perspective the need for clear instructions on data citation, derivative product, alignment of copyright, QA and provenance metadata were mentioned.

The analysis presented in this chapter aims to provide a mapping of the most prominent data stakeholders in the fisheries, marine and maritime sectors. Through this mapping, the licensing and other schemas for legal interoperability between the data sources represented, managed and owned by these stakeholders were identified and presented in order to support the profiling of the data interoperability ecosystem in the aforementioned sectors.

The results of this analysis so far showed that under the umbrella of various key data stakeholders there are a number of global research networks and data sources, providing access to data sets that are frequently used for building different data-powered services. Different datasets are covered by different licensing schemas.

The IODE Oceanographic Data Exchange Policy has been adopted by members of the IODE network and additional data providers, while the use of various CC licenses seems to be popular among other data providers. However, a high percentage of the existing data sources still use

¹⁰⁵ http://ec.europa.eu/maritimeaffairs/policy/marine_knowledge_2020/index_en.htm

¹⁰⁶ <http://www.emodnet.eu/>

¹⁰⁷ http://ec.europa.eu/fisheries/cfp/fishing_rules/data_collection/doc/swd-2015-118-final_en.pdf

¹⁰⁸ http://ec.europa.eu/fisheries/documentation/studies/retrospective-and-prospective-evaluation-on-common-fisheries-policy_en.pdf

proprietary licensing schemas and legal disclaimers of various types that are hard to be mapped to a universal schema, such as one based on the widely used CC. An effort has been made towards mapping existing licensing schemes in a CC-compatible table that will facilitate the mapping of these schemes to a CC-based licensing schema. However, due to the fact that this exercise took place in the context of desktop research, it is expected that more accurate information leading to more accurate mappings will be achieved through targeted interviews with the managers of the aforementioned data sources. In other cases, no information related to the policy applied on the use, reuse and sharing of content from the data sources.

The data analysis of the data sources studied in this report aims to identify the use of existing standards for the description of datasets and other research outcomes, so that they can be used for integrating information about the licensing schemas in existing or new metadata elements. However, it should be noted that not all stakeholder types identified in the deliverable D2.7 “Market Report on the Fishery and Marine Sciences Data Analysis Sector” were analysed in this version of the deliverable; in some cases this was due to the high number of data sources to be analysed (which was not the primary scope of this analysis) while in other cases, the identified stakeholders did not have a data portal/repository that could be analysed.

The summarised outcome of the consultations for EGI Engage are thus twofold; first a need for practical solutions where practitioners can find services that enable them to comply with legal interoperability needs from the data perspective, and another more community oriented process to assist in the definition and implementation of community wide data sharing arrangements.

EGI-Engage can contribute by enabling a framework where Legal interoperability is ‘just there’. Several key services that were mentioned during interviews that could be provided include:

1. **Reference data services** for Marine Fisheries Exploitation and Monitoring; especially for industrial fisheries, there is a need to access reference data that is of good enough quality to be used in on-board and on-shore systems. The FLUX example was often cited.
2. **Spatially explicit services;** Marine Fisheries Scientist that provide stock assessments have cross domain data needs. These currently are restricted since it is difficult to access, harmonise, and process the often large volumes of data. The lack of an EU directive, similar to INSPIRE for the access and control of cross domain processes was mentioned as an interesting case for a legally binding framework,
3. **Fisheries Traceability/Certification/Quality control Services;** the large number of data types from different stakeholders and the often sensitive and confidential nature of the data will make it difficult to design and implement a legal interoperability framework for data. Rather, security and confidentiality were cited as important issues for this community segment.
4. **Marine Environmental Data Services;** similar to point 2. above, the need for ‘harmonised’ data was often cited as a need, or rather on-line services to access large volumes of geospatially explicit quality data across boundaries; language, formats, storage location, metadata formats, release cycles, aggregation level were mentioned as ‘legal’ barriers impeding access and use.

EGI Engage can work with communities to develop services to comply with legal interoperability requirements:

In addition to the potentially rather generic services mentioned above, some members of the interview panel added opportunities where EGI Engage could develop specialised services for specific fisheries communities to support legal interoperability needs;

1. **Data services** for Fisheries impact and Ecosystem Studies, such as for Marine Protected Areas selection; these require a large quantity and variety of ad-hoc data;
2. **Secure services** for Maritime Surveillance (MCS) and Safety at Sea; a current challenge is processing data processing coming from different sources
3. **Compliance services** in Marine Fisheries Policy sector, as the EC is in the process of standardising fisheries reporting formats from all stakeholders (Member States, RFMO, etc.) through UN/CEFACT¹⁰⁹ standards and this could create business

The findings so far indicate that the existence of a legal interoperability framework that would allow or enhance the interoperability between the different data sources described in this document (but not limited to them). Any services to be provided by EGI-Engage to data stakeholders in the marine and fisheries context should ensure that this framework is taken into consideration so that legal aspects are automatically resolved – therefore the use, reuse and sharing of data will be allowed under the appropriate licensing.

¹⁰⁹ <http://www.unece.org/cefact>

5 Interview Analysis

The primary objective of the interviews was to gather information on the current legal barriers in sharing fishery and marine sciences datasets (raw and processed data) with other institutes and communities. The secondary objective was to collect recommendations to formulate a way forward. This analysis has been conducted through semi-structured interviews with representatives from relevant institutes. All interviewees are working for institutes (or departments of institutes) with as core activity data sharing; and are directly involved in the legal aspects of data sharing. The table below lists all interviewees.

| Name | Function | Institute |
|-----------------------|---|-----------------------------|
| Enrique Alonso García | Counsellor of State Head of the Division of Innovation & Science and Environmental, Rural & Marine Affaires | Consejo de Estado |
| Julien Barde | Researcher | IRD |
| Willi Egloff | Director and Legal expert | Plazi & EU BON |
| Aureliano Gentile | Information Manager | FAO-FIPS |
| Neil Holdsworth | Head of Data Centre | ICES & iMarine Board member |
| Johannes Keizer | Information Systems Officer | FAO-OPCC |
| John Latham | Head of Geospatial Data Modelling and Monitoring Unit | FAO |
| Nathalie Morcrette | Head of Intellectual Property and Contracts Unit | INRA Science & Impact |
| Nadia Nardi | Researcher and Project Manager | ENG |
| Marc Taconet | Chief | FAO & iMarine Board Chair |
| Paul Uhler | consultant in research data policy and management | N/A |
| Andres Vatter Rubio | Legal Officer | FAO-LEGN |
| Jeroen Steenbeek | Software Engineer | Ecopath |

Table 5: List of people interviewed in the context of this analysis

Lack of understanding of rules to share data

According to Julien Barde (Senior Researcher, IRD), a lack of awareness and understanding among data providers/users and legal advisors causes reluctance for data owners to share. Barde refers to an example where a research institute was willing to share data and had sufficient funding for legal advice. Unfortunately none of the consulted legal advisors were willing to accept the assignment due to the level of complexity. The same holds for data users, many do not know that

in some cases data providers are even obliged to provide the data. An increase in awareness of the rules will thus benefit both the data providers and users.

Aureliano Gentile (Information Manager, FAO-FIPS) has been directly involved in the instalment of the iMarine infrastructure. During this experience he found the lack of clarity on the use (or misuse) of derivative products a cause for retaining data. Gentile points out that the risk of reuse of data (derivate products) can be a reason for data owners to withhold data.

Pasquale Pagano (Researcher, ISTI Institute) highlights a misunderstanding on data sharing. Many research institutes are under the assumption that sharing is equivalent to placing in the public domain. This causes a needless reluctance to share data, since there are many other degrees of 'openness'.

Paul Uhler (independent consultant) states that a waiver of all rights is the optimal legal interoperability instrument from a strict legal perspective. This waiver will, of course, reduce the complexity of data sharing to a minimum and therefore be optimal. The second best option would be the frequently used, attribution only license. This type of license nonetheless knows many legal problems, 'but to always use a waiver when sharing data is of course a legal fiction' according to Uhler.

Uhler also discusses the lack of knowledge and capacity to understand legal agreements, which is conditional for data sharing. Uhler explains that he has experience with a variety of cases where e.g. a researcher signs a waiver of copyright protection without knowing that the researcher is not in the position to sign. Uhler concludes with a recommendation to increase knowledge and understanding among researchers in the legal aspects of data sharing, by normative as opposed to legal means. This normative approach could be easier to implement and pave the road for a legal solution.

Marc Taconet agrees and is convinced that the major legal barriers of data sharing in the fishery and marine sciences sector are a lack of knowledge and awareness of the rules and benefits of data sharing among both data users and providers.

Lack of resources major reasons for reluctance to share data

Johannes Keizer (Information Systems Officer, FAO-OPCC) has been involved in the development of many data sharing initiatives. Keizer states that major reasons for the reluctance to share data is a lack of resources such as time and knowledge. It frequently happens that data owners are willing to share data, but do not have (or claim not to have) sufficient resources to share this data. These resources are required for standardising the format and doing research on sharing conditions. A solution would be to make it easier for data providers to share data, Keizer mentions 'different shades of openness' as a solution to this problem. Keizer is convinced that there should be different standardised versions of open data.

Neil Holdsworth (Head of the Data Centre, ICES) argues that there still are some data owners who do not provide the data since they claim not to have resources to provide the data in the requested format. Holdsworth questions the validity of this argument since in most cases, since this only demands marginal resources. Also some specific legal clauses about the resolution of data, such as potential identification of vessels sometimes serve merely 'as an excuse' to not provide the data, says Holdsworth. Data should of course not reveal the identity of vessels which

should stay under privacy protection, but it is unfavorable when data owners use this argument unjustly.

Many do not realise that licensing is useless without copyright

Willi Egloff (Director and legal expert, PLAZI & EU BON) has a profound experience with the legal aspects of data sharing as an IPR lawyer. Egloff argues that researchers tend to have the idea they 'own data' unconditionally. But according to Egloff it happens frequently that data which is not copyrightable is licensed for use. Many research institutes do not realise that copyright protection is conditional for a license, a license after all only permits the use of some rights of the data to a licensee. This causes an unfavourable situation where data actually is open, but seems to be restricted due to the issuance of the license. When a data owner wants to restrict the use of the data without copyright, it should agree upon a contract with the data user.

Being first to publish results appears to be a major incentive to withhold data

Barde states that being first with publication of results based on research data for some institutes is a major reason to withhold data. Some institutes withhold research data at least until they have published results to prevent other institutes to gain reputation with their research data. From the moment of collecting primary data to publication of the results can take few years. The aforementioned few years are many considering the dynamic environment of the fishery and marine sciences sector.

Keizer also mentions the personal interest of research institutes as a major reasons to withhold data; he also states that some institutes use reasons such as 'lack of time and knowledge' to share data rather as 'excuses' to withhold data. Jeroen Steenbeek (Software Engineer at Ecopath) agrees with latter statement and states that data is sometimes 'hidden behind bureaucracy and personal interests'.

Citation should be standardised through a European citation Guide

Uhlir, Keizer, Holdsworth and Steenbeek state that a standardised citation guide would stimulate data sharing and thus legal interoperability. For Taconet, one critical aspect of legal interoperability concerns traceability of data related to the need to provide credit /visibility/metrics to ownership.

There is a need for clear rules through a top-down decision

Barde states that to foster data sharing, some institutes need to decide what should be confidential and what could be shared, and implement this through a top-down decision. A good system might be that funding agencies make clear what data sharing and formatting requirements are for the data they fund. Gentile is also convinced that the road towards open data, needs more top-down decisions on standards and rules of data sharing. However, Keizer argues that case studies should lead the way, this bottom-up approach with optimal examples from reality should lead the way for policy. Pagano explains that the pace of development of technology will always be higher than policy, therefore should entities directly involved in data management be involved in policy making. Taconet suggests that it would be desirable if all FAO data would be published under one common legal framework through e.g. terms of use. But there are major difficulties to develop this common legal framework in particular related to historical information and data held

in FAO systems and repositories, according to Taconet. FAO is in the position to take lead in this matter, and has proven to be capable with the implementation of the Vulnerable Marine Ecosystems (VME) databases and iMarine.

Public institutions that funded data collection should keep copyright

Uhlir recommends developing a top-down default rule of sharing data produced by public institutions. Barde explains that many data collection initiatives are EU funded, which means that the EU should keep the copyright on this data. This data is after all funded with public money, which means that this data should also be placed in the public domain. Pagano agrees with Barde and states that publicly funded data owners should be obliged to share their data through a top-down decision.

Uhlir points out that the top-down approach in the US with its statutory waiver of copyright protection for federal data, is optimal. This system automatically contributes federal data, i.e. data produced by the works of the U.S. federal government itself, and all edicts of any government regardless of level or whether or not foreign, to the public domain. The only drawback is that this increases the risk of commercial use without attribution which is deemed to be unfair. However, this could be prevented by making a distinction between commercial and non-commercial use.

Data sharing is all about trust, more attention should be given to normative standards

Holdsworth argues that data sharing is all about trust. Data providers are willing to provide data because they trust ICES. Transparency is key to establish this trust according to Holdsworth. ICES decided not to redistribute data, this is another factor that stimulates the feelings of trust. When data users want to access the original dataset, they will be redirected to the original data collector. ICES does not provide data for commercial use, due to the complexity of making the distinction between commercial and non-commercial use.

Data is more open under French law than under EU law

A new data protection law has just been adopted in France. The previous law distinguished research institutes from other data users, but this new law does not make a distinction. The Directive 2003/98/EC on the re-use of public sector information, otherwise known as the PSI Directive. This EU directive encourages EU member states to make as much public sector information available for re-use as possible. Previous to the creation of this directive this area was left to member states to regulate. This directive now provides a common legislative framework for this area. This is a good attempt to open data according to Morcrette, but it unfortunately excludes research institutes¹¹⁰. Data in France is a lot more open than under EU law says Morcrette, INRA is mainly subject to French law and so does not incur many problems with legal interoperability.

A standard for metadata should be adopted

¹¹⁰ Public sector information as defined by the PSI directive does not include documents held by educational and research establishments, such as schools, universities, archives, libraries and research facilities including, where relevant, organisations established for the transfer of research results.

Taconet discusses another major barrier for data sharing, which is a lack of metadata standards. Taconet states that in many cases a lack of sufficient metadata makes the data hard to ease. FAO could, also in this matter, take the lead and develop metadata standards.

Political aspects as a barrier for data sharing

Taconet argues that political reasons. The latter concerns countries that are incentivised to withhold fish catch data when e.g. vessels are infringing laws and illegal fishing. Taconet argues that *privacy* should be clearly defined, as he thinks this legitimate interest to withhold data is used merely as an excuse to withhold data.

Best practices

INSPIRE, DG Mare and Ecopath are best-practice cases according to Holdsworth. Major advantage is that datasets have to comply with metadata standards and existing format standards.

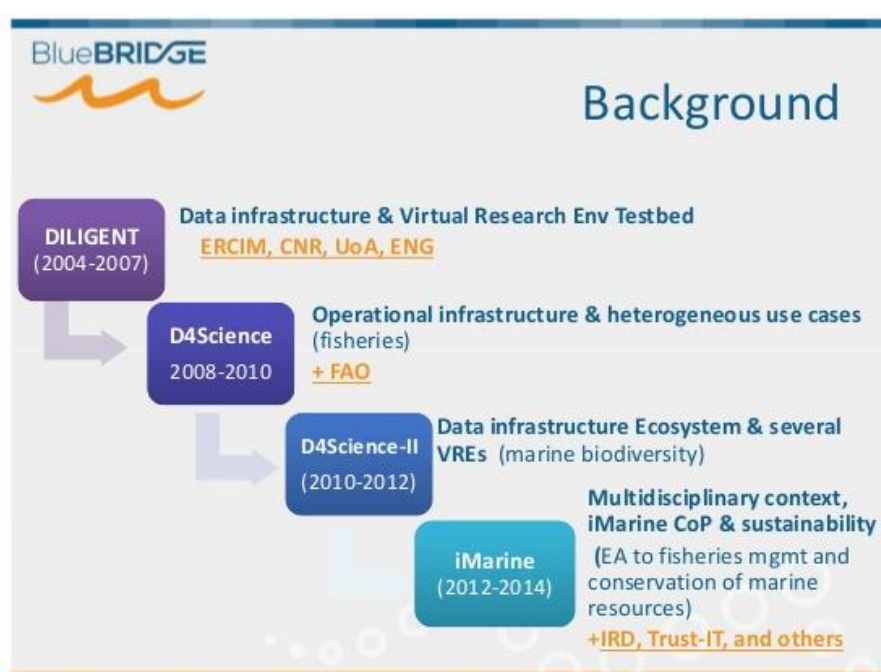
BlueBRIDGE operates as an aggregator of different sources and only provides the use of data instead of transferring ownership. The unique aspect is that the data is available to conduct calculations; but the data behind the parameters stays unshared. This method takes away many barriers to share data; but also increases the relevance of the trustworthiness of the data behind the parameters. Nonetheless, Pagano mentions the categorisation of data with metadata as an area of improvement for BlueBRIDGE. Pagano explains that there should be a standard for metadata that describes the data; once this standard is set, BlueBRIDGE should require all datasets to comply with metadata standards. Pagano also points out he doesn't understand why BlueBRIDGE supports SMEs, these commercial companies have customers and will make profits with this publicly funded infrastructure. Pagano therefore think it would be fair if commercial data users pay for BlueBRIDGE's services.

Taconet recommends

- (i) to implement more advanced licensing schemes, not only for data but also for its services and other derivative products;
- (ii) to engage iMarine in an active role in legal interoperability;
- (iii) to implement a system where data users can earn access to data by contributing data to this infrastructure;
- (iv) to improve attribution and the management of the *datatail* (where data originates and who should be attributed, i.e. traceability); and
- (v) to implement metadata standards through a top-down decision.

6 Use Case: BlueBRIDGE VRE in support of a Stock assessment Regional Database

This chapter illustrates the application of a data sharing framework to real life use-cases. BlueBRIDGE: Building Research environments fostering Innovation, Decision making, Governance and Education¹¹¹ is a project funded under the Horizon 2020 framework to provide innovative data services to scientists, researchers and data managers in the field of Blue Growth. BlueBRIDGE builds on existing EU and International e-Infrastructures providing capacity building in interdisciplinary research communities of scientists, data managers & educators in academic institutions. More specifically, the project aims to develop and exploit the existing iMarine e-Infrastructure data services for an ecosystem approach to fisheries.



BlueBRIDGE aims to validate a use case with a selected community of a regional database targeting fisheries' productivity. The selected use case, presented in the following sections, is currently being identified in the framework of the BlueBRIDGE H2020 project, and was only recently (January 2016) introduced to both the region (WECAFC) and the BlueBRIDGE project by the FAO of the UN. As of the writing of this chapter (February / March 2016) the activities for the development of this use case has not been started, and all mention of materials, methods, and collaboration in this chapter are related to real life examples, but not implemented. The validation of data sharing and legal framework has occurred at project organisation level; not at the community level.

¹¹¹ <http://www.bluebridge-vres.eu/>

BlueBRIDGE is a project funded under H2020 to **provide innovative data services to scientists, researchers and data managers in the field of Blue Growth and beyond**. The complete set of web-based data and computational resources offered by BlueBRIDGE will enable different stakeholders to **provide informed advice to competent authorities**.

The Western Central Atlantic Fishery Commission (WECAFC)¹¹² is NOT a partner in BlueBRIDGE, but works with FAO to identify the services and tools to offer through a VRE to its members and practitioners in the region.

The BlueBRIDGE project is driven by 14 European partners:

| # | BlueBRIDGE Partner | Country | Organisation Type |
|----|---|----------------|--|
| 1 | (Coordinator) CNR - CONSIGLIO NAZIONALE DELLE RICERCHE | Italy | Research institute |
| 2 | ERCIM – GEIE | France | Administration office |
| 3 | ENG - ENGINEERING - INGEGNERIA INFORMATICA SPA | Italy | Large company |
| 4 | UOA - NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS | Greece | University |
| 5 | FAO - FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS | Italy | Intergovernmental organisation |
| 6 | ICES - INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA | Denmark | Blue Domain related scientific institution |
| 7 | FORTH - FOUNDATION FOR RESEARCH AND TECHNOLOGY HELLAS | Greece | Research institute |
| 8 | IRD - INSTITUT DE RECHERCHE POUR LE DEVELOPPMENT | France | Blue Domain related scientific institution |
| 9 | TRUST-IT - TRUST-IT SERVICES LTD | United Kingdom | SME |
| 10 | I2S - OLOKLIROMENA PLIROFORIAKA SISTIMATA | Greece | SME |
| 11 | CITE - COMMUNICATION & INFORMATION TECHNOLOGIES EXPERTS S.A. OF CONSULTING AND DEVELOPMENT SERVICES | Greece | SME |
| 12 | CLS - COLLECTE LOCALISATION SATELLITES SA | France | Large company |
| 13 | GRID ARENDAL | Norway | Large company |
| 14 | PMBret - ASSOCIATION POLE MER BRETAGNE | France | Business & Innovation Cluster |

The BlueBRIDGE goal is:

to support capacity building in research communities involved in resource exploitation, environmental and ecosystem management with the aim to provide informed advice to competent authorities and to enlarge the spectrum of growth opportunities as addressed by the Blue Growth Societal Challenge.

¹¹² <http://www.fao.org/fishery/rfb/wecafc/en>

What are Virtual Research Environments?

The BlueBRIDGE project develops support services in an existing e-Infrastructure, and combines these in Virtual Research Environments (VREs).

A Virtual Research Environment is a **system that provides researchers and research teams, educators, SMEs, and any other type of user, from different disciplines, institutions or even countries, with a web-based set of facilities including services, data and computational facilities.** These facilities are **tailored to serve the needs of a specific Community of Practice, that is a set of individuals who decide to “virtually”** (they do not necessarily need to be formal structures such as departments or project teams) **connect to solve a specific problem.**

Rather than presenting isolated services, VRE's allow to compose an organised and managed set of facilities that not only combine data, processing and users, but explicitly also includes data sharing policies. Through a focus on metadata management, the VRE offers handles for refined implementation of fine-grained policies that include attribution, copyright management, citation policy, provenance management, and licensing of single and derivative products. The VRE terms of use are therefore composed according to the combined data and services selected at the time of VRE definition. Raw data are then licensed according to the license expressed by the data owner/custodian and expressed at time of registration of the data content to the infrastructure. Derived data products instead have to be licensed with a license compatible and legally interoperable with the one associated with the primary data. As explained in Section 4.2 this task is not always fully managed by the infrastructure enabling technology and cannot be ensured always. A set of licenses for derived data products (according to the license associated with the primary data, e.g. either CC BY-SA or CC BY-NC-SA license,) is proposed to the users together with all the information about the licenses of the primary data exploited. It is then under the responsibility of a single user, as expressed in the VRE terms of use, to select and confirm the license to associate with any produced derived data.





The use of an infrastructure for controlled data sharing opens new perspectives for data sharing policies that use services to track compliance. E.g. a data owner may wish to offer services on top a dataset for a specific audience that includes a download quota, aggregation levels for display, open data for computations but not for download (share the results, keep the data), and proper citations and back-links to exposed data. These services do not always exist, and BlueBRIDGE will develop them but the aggregation levels for display.

The BlueBRIDGE services are the customisable components that organisations such as WECAFC can use to define a comprehensive data policy. These are then combined in a VRE that can be managed by the VRE owner for the time and community that has been made available to that VRE.

In this use case, we will validate the VRE capabilities to offer services related to data sharing and legal interoperability. For other types of interoperability, as semantic and technological, appropriate notes will be added. To validate if the VRE's meet legal interoperability standards, we apply a simple framework where for each 'managed' asset (data or computing) we question if any questions remain related to:

1. Intellectual property rights;
 - a. Copyright;
 - b. Database protection right(sui generis);
2. National security laws;
3. Confidentiality laws and policies;
4. Protection of endangered species;
5. Individual contracts or agreements.

The table below summarises the relevant VREs that are relevant to the WECAFC RDB:

| BlueBRIDGE VREs | | Related objectives |
|---|---|--------------------|
|  | Stock Assessment VRE: providing an on-line environment for Stock Assessment for Blue Growth practitioners, with the long term strategy to produce evidence based understanding of the status of marine fisheries. | Blue Assessment |
|  | Global Record of Stocks and Fisheries VRE providing an on-line knowledge base on the Global Record of Stocks and Fisheries for a Blue Growth audience of ecologists, resource managers, market parties and the general public with the long term objective to provide evidence based information on the status of marine stocks and fisheries and promote responsible consumption. | Blue Assessment |
|  | Protected Area Impact Maps VRE providing scientists with an integrated environment supporting the efficient and effective production of maps of vegetation types and human impacts on them and enabling ecosystem degradation analysis. | Blue Environment |
|  | Knowledge Bridging VRE provides services to educators for setting up cost-effective training environments for (real) data analysis and computational resources in a short time. | Blue Skills |

To develop these services, the existing **D4Science infrastructure** will be used, while the iMarine initiative will organise the wider community by ensuring sustainability to the VRE's. The D4Science infrastructure will also transparently bridge between computing and data infrastructures: all the activities in the project related to the development and deployment of **service and resource commons** to facilitate the exploitation of existing infrastructure resources are referred in the project as Blue Commons.

The BlueBRIDGE RDB VRE will be made available via the iMarine Gateway¹¹³.

6.1 WECAFC Stock Assessment VRE

The stock assessment VRE for WECAFC will be a specialisation of a broader Stock Assessment VRE. It will **provide an on-line environment for Stock Assessment, with the long term strategy to produce evidence based understanding of the status of marine fisheries.**

It will rely on existing infrastructure services (in bold) for:

- data sharing through a **WorkSpace** and an **SDMX registry** for statistical data (timeseries mainly);

¹¹³ i-marine.d4science.org

- data processing with R and Java through a **Statistical Manager** service;
- data harmonisation with **SpeciesDiscoveryService** and **BiOnym** for name disambiguation;
- data dissemination and provenance metadata management through a **WorkSpace** and **TabularDataService**;
- **identity management, authorisation, and accounting** services

6.1.1.1 *Specific challenges*

WECAFC data are ownership of the members, and this will have to be reflected in all data and derivative products. The ownership of any property rights is not in any way transferred to the Infrastructure.

A legal framework in the region for data sharing does not exist, and before collecting data in the infrastructure, all contributors will require assistance in the entire process of moving their data to an infrastructure.

None of the owners in the region (outside the EU) is likely to have knowledge of hosting data and data processes on a cloud infrastructure with a focus on Europe. This carries risks of lengthy legal procedures to obtain release certificates and shared ownership and management responsibilities in an infrastructure setting.

Information on fish population (e.g. growth, natural mortality, and recruitment), fisheries (e.g. selectivity and the environment) is used in fisheries stock assessment and management. Obtaining this information is difficult, as it is currently stored in institutes who do not have a clear data policy or mandate to develop such a policy.

When processing data, a similar lack of data sharing policies will affect the processing stage. If no quality indicators are available for input data, a policy on reliability and precision cannot be applied.

Once data are processed, preventing that incorrect assumptions are made with a substantial impact on stock assessment results and resulting management advice. The main challenge is to provide a dedicated environment for stock assessment, including

- Access to relevant data;
- Access to a range of related algorithms and models;
- Availability of storage and dissemination tools.

The complexity of stock assessment, and the widely varying quality and availability of data and algorithms, is particularly suited to a transparent and collaborative approach. The environment offers a flexible set of algorithms that are as easily discovered as shared, and results are easily published.

For individual scientist with sophisticated models the computing capacity offered through this VRE promises to boost productivity and, when properly embedded in the underlying infrastructure, citable and reproducible scientific experiments.

6.1.1.2 *Functionality validation related to services for data sharing and legal interoperability*

Related to IPR management

- Data sharing policy; Workspace transparent to data contributors
- Data analysis and Modelling; Workspace and Statistical Manager support rich metadata records related to ownership and applied process tracking in standard format and auditing tools
- Data repositories (SDMX) support refined policies on access, download and upload
- Data provenance and attribution management, also for derivative products are automatically generated in Prov-O format and then made accessible through the workspace.

Related to national security laws

- Not relevant

Confidentiality laws and policies

- The workspace offers a rich set of facilities to manage data access rights.

Protection of endangered species / sites

- Data aggregation facilities are available

Individual contracts or agreements

- To be determined is needed
- No specific facilities are available (or requested)

6.2 The Global Record of Stocks and Fisheries VRE and WECAFC data

The Global Record of Stocks and Fisheries VRE provides an **on-line knowledge base on the Global Record of Stocks and Fisheries for a Blue Growth audience** of ecologists, resource managers, market parties and the general public with the long term objective to provide evidence based information on the status of marine stocks and fisheries and promote responsible consumption.

The goal is to set up a **global platform**:

- For compiling and sharing stock assessment and management data for all of the world's fisheries
- To facilitate access to information on status & trends of stocks and fisheries
- To offer services to compute regional/global state of stocks indicators
- To advocate the improvement of data collection and governance in data poor contexts

Especially the last point is of relevance for WECAFC, who miss the resources to establish such a featureset in their own region. Through BlueBRIDGE pooling of resources, the data collated by WECAFC can be made available to a global audience through a single point of service, and become part of a global commons of fisheries data. This can be a very cost effective data sharing for both data producer and consumer.

However, this data sharing across VRE's comes with not a few data sharing and legal interoperability challenges.

6.2.1.1 Specific challenges for WECAFC

Information on the status and trends of stocks (the exploitation level) and fisheries (the ecological and economic sustainability) depends on a multi-stakeholder analysis of information to produce a shared opinion. This opinion is mostly based on stock assessment information (the previous sub-case) and an assessment of the management capacity to ensure future sustainability of the fish and the fishery.

The challenge with WECAFC assessments is that it is not uncommon that there are differences in opinion between different assessment groups, and a single report a stock cannot be always produced.

From a data sharing perspective, how can we then ensure that WECAFC data are available for review, but are not widely distributed? The main challenge is to provide an effective communication mechanism on WECAFC stocks and fisheries that builds on existing information resources, but renders these:

- Accessible through web-services and/or a web-interface in a Knowledge Base (the GRSF KB);
- Clear and easy to understand for users (e.g. using traffic lights, or map viewers);
- Manageable at authorisation levels, e.g. WECAFC data managers and generic users will have different data views, and release control is in the hand of WECAFC assigned staff for sharing and dissemination;
- Transparent through traceable and well described provenance and ownership metadata;
- Communicative; the KB is integrated with query and display features, including a map-interface;
- Dynamic; by delivering tools and services to collect and merge data with the GRSF KB.

6.2.2 VRE specific data policy and legal challenges, functionalities

6.2.2.1 Functionality validation related to services for data sharing and legal interoperability

Related to IPR management

- Data integration of WECAFC information sources can use a set of minimal data requirements
- Data dissemination and publication is covered in GRSF KB
- Data provenance management is provided through GRSF KB, but will be further developed
- Data merging and mapping tools will be developed in BlueBRIDGE
- Data repositories (LOD) are available in GRSF and managed in an infrastructure

Related to national security laws

- Not relevant

Confidentiality laws and policies

- The GRSF will have to offer facilities to manage data access rights.

Protection of endangered species / sites

- Not applicable

Individual contracts or agreements

- To be determined is needed
- No specific facilities are available (or requested)

6.2.3 Protected Area Impact Maps VRE and WECAFC spatial data

The Protected Area Impact Maps VRE will provide scientists with an environment for the production of maps of vegetation types and human impacts on them and enabling ecosystem degradation analysis. The aim is to facilitate assessment of protected areas (support Natura 2000 and Aichi target 11), and develop management responses by linking ecosystems with potential impacts. It thus deals with spatial data, and implements policies for their management and sharing.

Some of the datasets prepared in the WECAFC scenario will be useful to this VRE, but a data policy is required that ensures that only data that has been released for publication by the owners arrives in this VRE for public use. Other data products may be produced in other workflows, and an agreement has to be established with the data owners that recognises their IPR and other data sharing arrangements. A further characteristic of this VRE is that FAO liaised with JRC and CLME+, who are not partners in BlueBRIDGE, for development and exploitation of the VRE. This requires a carefully balanced policy that ensures that spatial data provided by the WECAFC partners is used correctly and transcends the boundaries of a specific infrastructure; it must be global in scope of its definitions and applicability. This VRE scenario presents an opportunity for the development of an infrastructure based socio-economic and environmental impact assessment of MPAs in the WECAFC area supported by a global data policy for spatial data.

6.2.4 VRE specific data policy and legal challenges, functionalities related to WECAFC potential products

6.2.4.1 *Functionality validation related to services for data sharing and legal interoperability*

Related to IPR management

- OGC services integrated in the infrastructure can provide IPR management for WECAFC data
- OGC services are available to cover IPR issues related to spatial data dissemination and publication
- Data provenance management is provided through GRSF KB, but will be further developed
- Spatial data merging and mapping tools will be further developed in BlueBRIDGE
- Spatial Data repositories are available in and managed through an infrastructure
- BlueBRIDGE will have to deliver a service to manage citation of derivative maps (based on more than one source)

Related to national security laws

- Not relevant

Confidentiality laws and policies

- OGC services provide out-of-the-box facilities to manage data access rights and policies.
- No relevant laws have been mentioned in the context of spatial data management

Protection of endangered species / sites

- Not applicable

Individual contracts or agreements

- To be determined if needed

7 Findings and Opportunities

The work described in this document aimed at mapping the data stakeholders at a global level in order to identify the current status of the legal interoperability framework, focusing on the policies that apply regarding the use, reuse and sharing. The desktop research, consisting of web search for related content combined with existing resources on the topic and enriched/validated with the input elicited from domain and other experts, has highlighted the issue of the lack of a common legal framework that will allow the seamless interoperability between different data sources in the marine, fisheries and maritime sector.

On top of that, the following findings have been elicited from the research described in this report:

1. **Lack of understanding of rules to share data.** Transparency could facilitate the decision-making processes for business considering how to handle transborder data flows. Transparency could be increased by making all relevant texts of national laws and regulation on data protection, particularly on transborder data flows, available in different languages. Reusers of data need to clearly understand what they can (or cannot) do, without asking their lawyers and, ideally, without reading too many licenses. Lack of the education on the benefits and the application of open licenses over their data is one of the causes for this issue.
2. **Lack of resources is one of the major reasons for reluctance to share data.** Publishing data online and ensuring that the author/creator still maintains his/her rights on the data (e.g. ownership, attribution) is a time-consuming process which requires research for understanding and selecting the most appropriate licensing schema. For example a researcher who wishes to publish his data needs time not only to be informed about the appropriate licenses but also to find out how they will be applied over his dataset and also have the technical expertise to find an appropriate data repository, describe his dataset with metadata and publish it. In case of large datasets, a fast network/internet connection is also needed for uploading these datasets to the data portal/server.
3. **Being first to publish results appears to be a major incentive to withhold data.** Despite the existence and application of standard licensing schemas for stating the ownership of data and the terms of use, reuse and sharing, data producers and researchers still reluctant to be the first to publish their data.
4. **Citation should be standardised through a European citation Guide.** The existence of various citation formats is confusing to people who wish to publish their data online and get the credit for it. Data producers need to feel assured that their data is properly described and cited so that they can be referenced as the owners or producers of the data.
5. **There is a need for clear rules through a top-down decision.**
6. **Public institutions that funded data collection should keep copyright.** Despite the fact that we mostly discuss about open data, the funders of the research leading to the production of research data (such as research and higher education institutes, funding bodies etc.) in some cases need to maintain the copyright over the data.
7. Data sharing is all about trust, more attention should be given to normative standards

8. **As long as relevant government measures are not adopted and enforced in all jurisdictions, this analysis recommends the use of common-use licenses or waivers of intellectual property rights on a voluntary basis for data, metadata and the reuse of this data.** As expected, it is easier to apply the appropriate licenses when the mandate comes from a national, international (e.g. EU / Horizon 2020) or organisational (e.g. funder's) level. This allows the automatic and horizontal application of the most appropriate licensing schemas on the research data produced under certain circumstances (e.g. with a funder's, a state's or a project's funds).
9. **Provide free open data to data users in data-poor areas such as developing countries where they cannot afford the data.** Despite the fact that developing countries are producing data, it is not always the case that they have access to them. Lack of infrastructure (e.g. data platform), expertise and funds required for this effort are among the top reasons for this. By offering open data to developing countries (and other data-poor countries as well), members of the local communities will have the opportunity to develop innovative products based on them, as well as to improve their activities (e.g. production, marketing etc.).
10. **There should be a quality certificate for data.** Taking into consideration existing efforts in the assessment of the quality of data (such as the Open Data Certificate¹¹⁴ of the Open Data Institute and the 5-star Open Data¹¹⁵), a mechanism for the assessment of data available in the fisheries, marine and maritime sciences should be developed or adapted and adopted (in the case of existing ones). This will allow the identification of high quality datasets in terms e.g. of completeness and content, so that they can be further promoted and reused in various applications.
11. Pagano also mentions a quality certificate (gold, silver etc.) for data. When the quality of a dataset is easy to assess, it is easier to only work with the data online without transferring ownership. When you can be sure that the data is of a certain quality.

The following table summarises the findings of this work as well as a set of proposed actions that may be taken towards addressing the corresponding legal interoperability issues.

| Problem | Action | Responsible |
|--|---|-----------------------------------|
| Lack of understanding of rules to share data | Various types of stakeholders in the research data value chain need to be better educated on the legal aspects of data sharing. | Open data experts, legal officers |
| Lack of resources major reasons for reluctance to share data | The requirements of data producers, managers and other types of data stakeholders should be identified and the necessary resources should be made available to them. Better education on data sharing will eliminate issues related to e.g. lack of time, as applying the appropriate license will become a standard procedure. | Depends on each case |
| Being first to publish results appears to be a major incentive | By applying the appropriate license on research outcomes, researchers and other data producers maintain ownership | Data managers, legal officers |

¹¹⁴ <https://certificates.theodi.org/en/>

¹¹⁵ <http://5stardata.info/en/>

| | | |
|---|--|--|
| to withhold data | of their data and at the same time they are acknowledged each time this data is used. | |
| Citation should be standardised through a European citation Guide | Despite the continuous efforts of the EC towards applying open access and open data policies over the EU-funded research projects, still public data are not openly shared. An EU-proposed citation guide would set a standard to be applied by researchers and other data producers over their data | High level policy makers, EU officers |
| There is a need for clear rules through a top-down decision | By engaging stakeholders at an organisational or network level in clearly defining and applying data sharing policies over their data, all related data sources will be obliged to comply with the proposed decisions. | High level policy makers, network coordinators |
| Public institutions that funded data collection should keep copyright | The concept of data sharing and the available options (including the ones that allow data producers and managers to retain copyright over such data) should be better explained. | Data management officers at national level |
| Data sharing is all about trust, more attention should be given to normative standards | Data stakeholders should be better educated on the application and use of standard licensing schemas for data sharing. | Open data experts, legal officers |
| Provide free open data to data users in data-poor areas such as developing countries where they cannot afford the data. | Developing countries may act as data producers; at the same time, they may lack the capacity (technical, financial etc.) required for the storage, management and exploitation of data, as well as the reuse of data from other sources. Special provision for facilitating access to the necessary resources as well as actual data should be made for these countries. | Data producers & managers, high level policy makers, legal officers and experts. |
| There should be a quality certificate for data | The quality of data needs to be evaluated and measured with qualitative and quantitative metrics. This will allow those responsible to provide a quantitative score representing the quality of data. | Data certification bodies, data experts |

Another interesting finding is that data properly and openly licensed through a license allowing their use (even for non-commercial purposes) and sharing were exploited and used for the development of related data-powered services. Such data from the fishery and marine context were used (as described in the previous chapters of this document) for the enrichment of related maps, the development of models of interest to the fishery and marine researchers as well as in various other applications. By ensuring that such data are properly acknowledged, the necessary citation was provided as a template in several cases and this led to the increase of the use of the data in a trackable mean.

Last but not least, this chapter discusses clear, final conclusions, which can easily be translated into a way forward. The conclusions are presented in a brief and comprehensive manner in the table below. The findings are based on the 14 interviews conducted and on insights from the author of this document. The findings of this legal analysis are based on the findings from the *Market report on the Fishery and Marine Sciences Data Analysis Sector (D2.7)*.

| Findings | Proposed EGI actions |
|---|--|
| The Market Analysis claims that legal and governance (legal restriction to use cloud-based services, lack of clear roles of ownership) is the top challenge/obstacle recognised for cloud-based services at 57 percent. | To propose a legal and governance framework, taking into consideration the actual needs of all stakeholders, existing efforts and standards as well as restrictions posed by legal entities (e.g. governmental and institutional policies) |
| Fishery exploitation is mainly private, but monitoring and management is mainly public. This means that data collection is mostly publicly funded. Public institutions that funded data collection should keep copyright. | To envisage the rights of data producers and ensure that data produced through public funds will be openly shared where possible. |
| The market analysis pointed out that the processing of data coming from different sources is a challenge. All respondents of the survey presented in the Market Analysis need access to external data and many from other domains. This lack of legal interoperability has thus also shown to exert a negative effect on business opportunities in this sector. | To promote the alignment of data use and sharing policies and encourage the adoption of existing standards. To promote the mapping of existing proprietary licensing schemas to existing standards in order to facilitate the interoperability between different data sources. |
| Regarding fisheries data, there are as many formats as countries/regional institutions/international organisations. Standardisation of data and legal frameworks to share data are essential. | To promote the use of existing standards for the description of data (i.e. metadata standards) and the use of standard data formats in order to enhance interoperability. To take into consideration existing standards. |
| Fisheries data worked with is increasingly global, international legal interoperability is therefore increasing | To ensure that the standards to be adopted will be valid at a global level, taking into consideration the interoperability with data stakeholders and sources outside EU. |
| Lack of understanding of rules to share data, the Market Analysis also argues that currently few information on regulations of data sharing is transferred to scientists who work on data. | To ensure that researchers, as data producers, are properly educated on the importance and use of licenses over their data. |
| Lack of resources major reasons for reluctance to share data | To support data providers by improving access to the necessary resources that will facilitate the sharing of data; e.g. to educate data providers on the importance of data licensing and sharing of data. To reduce the effort needed by data providers in properly licensing their data e.g. by adopting existing licenses to be automatically applied to data at an organisational level. |
| Being first to publish results appears to be a major incentive to withhold data | To better educate researchers and other data providers on the rights they have on their produced data and the benefits of |

| | |
|---|--|
| | applying the appropriate license on research outcomes. |
| Citation should be standardised through a European citation Guide | To consider an EU-proposed citation guide that would set a standard to be applied by researchers and other data producers over their data at EU level. To ensure that existing efforts are taken into consideration. |
| There is a need for clear rules through a top-down decision | To engage stakeholders at an organisational or network level in clearly defining and applying data sharing policies over their data. In this way, all related data sources will be obliged to comply with the proposed approaches. |
| The CWP and FIRMS metadata standards are insufficient. | To identify the needs of the people working with metadata standards in specific cases (such as CWP and FIRMS) and ensure that they will have access to all resources needed for adapting a better metadata standard, which will ensure technical and legal interoperability. |
| Data sharing is all about trust, more attention should be given to normative standards | To ensure the adaptation and use of existing standards among various types of data stakeholders in the fishery and marine context. |
| Provide free open data to data users in data-poor areas such as developing countries where they cannot afford the data. | To ensure that developing countries have access to the necessary resources (e.g. technical and financial, expertise etc.) as well as access to actual data that may be of use to them. |
| There should be a quality certificate for data | To define a process for the evaluation and assessment of the quality of data with qualitative and quantitative metrics. This will allow those responsible to provide a quantitative score representing the quality of data. |

The findings of this work are expected to drive the next steps in enhancing legal interoperability between the various data sources and stakeholders in the fishery and marine sector. The first step towards an integrated effort for enhancing the overall interoperability between the aforementioned actors is to remove barriers related to legal interoperability. This will make data more easily available and reusable therefore additional work and will enhance other types of interoperability (e.g. semantic and technical).

8 Conclusions

Rapid technological developments challenge data sharing policies and legal frameworks in the fishery and marine sciences sector. The existing normative rules of data sharing are insufficient to sustain the high pace of technological developments. There seems to be a gap between a normative set of rules (attribution to avoid plagiarism) on the one hand, and policies and legal frameworks (licenses, contracts, etc.) on the other, for individuals working with research data. The rapid technological developments thus increase the need for a clear and crystallised set of rules for data sharing.

This document analysed the legal barriers to share data, and revealed that there is a need for a clear direction in this discussion. This direction initiated with a definition of *legal interoperability*. The legal interoperability instruments have a pivotal role in stimulating data sharing. Waiving all copyright protection is, from a legal perspective, the most favourable option to stimulate data sharing. But there are many incentives discussed in this document to withhold data. The legal interoperability instruments satisfy these incentives. But a variety of legal interoperability instruments poses a negative effect on data sharing. When aggregating data from multiple data sources with custom agreements or common-use licenses, the data user will have to comply with the accumulation of all conditions and restrictions. The author learned writing this document that in some cases concerns more than 70 different agreements or common-use licenses. This obviously hampers data sharing and reveals a need for standardised common-use licenses such as CC licenses.

There are promising initiatives currently working on the (legal) interoperability issues such as INSPIRE, RDA-CODATA's Interest Group on Legal Interoperability and GEO.

Based on the findings of the legal analysis of the fishery and marine data sources which was described earlier in this document, a step towards enhancing (and ensuring) legal interoperability between these data sources is the definition of a data access, use and sharing policy that will meet the requirements of these data sources. In case an existing standard licensing schema is used (such as a CC license), effort should be made towards the mapping of existing legal requirements regarding the use and sharing of data to this license. This will require consultation meetings with various types of stakeholders that are involved in data production and management in the marine and fisheries sectors, such as data producers (e.g. researchers), legal officers, repository managers, operators of aggregators, data managers, librarians and any other type of stakeholder identified. This will eventually allow the replacement of custom, hard to use data policies with existing standards that will significantly enhance the interoperability of data between different systems, platforms and data sources.

Licensing schemas based on the CC licenses are considered the first step towards ensuring the necessary legal interoperability between different data sources in the fishery and marine sciences. They provide the flexibility to meet different licensing options and provide a number of advantages, such as the machine readable format of a license, which significantly facilitates the retrieval of CC-licensed content through queries performed with search engines. The use of existing standards in the fishery and marine ecosystem of data sources should also be considered

as well as existing efforts in the broader context - such as the one in the geospatial sector which has a tradition in developing and adopting standards for enhancing the interoperability between various data sources in the specific domain.

The brief analysis of the data available through the various data sources in the fishery and marine context aimed to showcase the complexity of the specific ecosystem in terms of data types, formats and interoperability options. This analysis highlights the need for licensing schemas to cover not only a wide variety of data types, formats and services built on them, but also to address the legal issues raised when e.g. data are aggregated from their original sources (e.g. repositories) and then made available through a different license by the aggregator.

As expected, only open data or at least data with the appropriate licensing can be reused for building added value services on top of them, thus enhancing their value and exploiting their full potential. Such data can also be widely distributed among interested stakeholders through the use of interoperability standards (e.g. technical interoperability options like OAI-PMH and REST APIs, among others), and also combined with related data through their exposure as linked open data (referring to the semantic interoperability).

Additional aspects on the conclusions extracted in the context of this work are presented in the following sections.

Relevance and applicability of legal interoperability for data driven developments

Data-driven innovation will benefit from the resolution of issues related to the use, reuse and sharing of data from various sources in the marine and fisheries contexts. This applies not only in cases where the existing license restricts the use of data (in some cases for no obvious reason, e.g. due to the selection of an incorrect license) but also in cases where no license has been applied to data, making its exploitation impossible or risky (e.g. in the case of commercial use of copyrighted data). For example, the work of data-powered SMEs that develop innovative applications based on open data will be significantly facilitated by setting a clear licensing schema over existing datasets that can be used for commercial purposes. In the case of researchers, they will also benefit from the use of previously published data that can be used for non-commercial purposes, as long as the appropriate licensing has been applied to this data. What is important in all cases is the existence and application of a licensing schema which clearly defines the allowed uses of the specific datasets; based on that, data-powered applications and services can be built respecting any restriction described in the corresponding license.

Obstacles and potential exist in relevant markets / scenarios / studies

Various types of stakeholders in the data market (and not limited to it) will benefit from a clear and standard legal framework applied over marine and fisheries' datasets and services. For example, individual developers will be able to develop innovative data-powered applications either in the context of hackathons or even at a commercial level. This is expected to motivate both individuals and SMEs working with open data applications for commercial or research purposes.

The reuse of data for research and academic purposes will also benefit from that; researchers will have access to datasets and services that can be used for non-commercial purposes in order to support their research and based on existing data, to move their research to the next level. Spin-

offs deriving from research and educational institutes will have access to such datasets and will be in the position to develop commercial products based on open data (available for commercial purposes).

Research institutes at a national and international level will benefit from combining their data with other freely and openly available data from other sources and their integration in their services. At the same time, aggregators will be able to harvest data and metadata, which are appropriately licensed, from various sources and enhance their discoverability and retrieval by researchers and other stakeholders at a global level, having significant impact on the work of researchers that depend on this data. Publishers will be able to combine the research publications of their journals with related datasets from previous studies, thus enriching the added value of their products. This is expected to have significant impact on the value of the research publications and enhance the exploitation of datasets from researchers working on the same topic.

The outcomes of this deliverable should be studied in conjunction with the corresponding outcomes of the deliverable D2.7 “Market Report on the Fishery and Marine Sciences Data Analysis Sector” in order to obtain a complete view of the fishery and marine research data ecosystem.

9 Bibliography

- Assante M., Baldacci M. B., Candela L., Castelli D., Tani A. (2015) Research data publishing in scientific data repositories: a survey. ISTI Technical Report.
- Assante, M., Candela, L., Castelli, D., Tani A. (2016) Are Scientific Data Repositories coping with Research Data Publishing? Data Science Journal (under review)
- Chisum, Donald S., et al.(2011): Understanding intellectual property law. LexisNexis.
- de Cruz, P. (1999). "Comparative Law in a Changing World". London: Cavendish Publishing.
- Doldirina, C. (2013): Interpretation of the "full and open" access to and use of (geographic) data: existing approaches, Living paper of the GEO Data Sharing Working Group.
- Egloff, W. & Agosti, D. (2014): Case Study: Plazi Treatment Repository. RDA-CODATA Interest Group on Legal Interoperability of Research Data
- Fletcher, W.J., Chesson, J. Fisher, M., Sainsbury K.J., Hundloe, T. Smith A.D.M., and B. Whitworth (2002): National ESD reporting framework for Australian fisheries: The "How To" guide for wild capture fisheries. FRDC Project 2000/145. Canberra, Australia
- Gasser, U. (2015): Interoperability in the Digital Ecosystem, *The Berkman Centre for Internet & Society at Harvard University*, Research Publication No. 2015-13
- Global Aquaculture Production Fishery Statistical Collections, FAO, Rome. Retrieved 2 October 2011
- Janssen, K. (2011): Legal Interoperability – barriers to the harmonization of licenses, Share PSI workshop Position Paper. Leuven, Belgium
- Lammerant. H, Galetta A, De Hert P. (2014) *Legal issues in big data in Anna Donovan, Rachel Finn & Kush Wadhwa (eds.), Report on legal, economic, social, ethical and political issues regarding big data. BYTE Deliverable D2.1, 35-58. BYTE Deliverable D2.1: Report on legal, economic, social, ethic ed. Unknown, 2014. 24 p.*
- Legal Options for the Exchange of Data through the GEOSS Data-CORE*, GEOSS legal Interoperability Sub-Group Data Sharing Task Force (20110
- Magron, F. (2015): *Towards a data sharing policy for survey and monitoring data collected and/or stored by the SPC Coastal Fisheries Programme on behalf of member countries*, SPC Coastal Fisheries Science and Management Section, Working Paper 14
- Miller, A D. (2007): "The reasonable man and other legal standards. California Institute of Technology, Social Science Working Paper No. 1277
- Miller, P. (2000): "Interoperability. What Is It and Why Should I Want It?", Hull, Ariadne Issue 24. Retrieved at 18-01-2016 at <http://www.ariadne.ac.uk/issue24/interoperability/>
- Morano, F. (2013): Legal Interoperability: making Open Government Data compatible with business and communities. JLIS.IT, vol. 4 n. 1, pp. 441-452

Nardi, N. (2016): Market Report on the Fishery and Marine Sciences Data Analysis Sector, EGI-Engage

OECD Principles and Guidelines for Access to Research Data from Public Funding, OECD (2007)

Oppenheim, C., Korn, N., (2011): Licensing Open Data: A Practical Guide, June 2011 version 2.01

Palfrey, J.G. and Gasser, U.(1999): Interop: The promise and perils of highly interconnected systems. Basic Books, 2012.

RDA-CODATA Interest Group on Legal Interoperability of Research Data, iMarine (2014)

Weber, R. (2009): "Mapping and Structuring International Financial Regulation – A theoretical Approach." *European Banking Law Review* 20 (5): 651-88

Weber, R.H. (2014): Legal Interoperability as a Tool for Combatting Fragmentation, Global Commission on Internet Governance Paper Series: NO. 4.

White Paper: mechanisms to share data as part of GEOSS Data-CORE, GEO (2013)

de Witt, S. & Brzezniak, M. (2015): *D7.4: Recommendations for techniques and policies to enable optimized data access*