

**EGI-Engage**

D2.6 – Report data sharing policies and legal framework in fishery and marine sciences

|  |  |
| --- | --- |
| **Date** | 29 February 2016 |
| **Activity** | NA2.3 |
| **Lead Partner** | FAO |
| **Document Status** | DRAFT |
| **Document Link** | https://documents.egi.eu/document/2699 |

Document describes constraints and requirements given by IPRs and overall legal framework 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.

**COPYRIGHT NOTICE**



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

**DELIVERY SLIP**

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***Name*** | ***Partner/Activity*** | ***Date*** |
| **From:** | Eise van Maanen (FAO), Anton Ellenbroek (FAO) | FAO/NA2.3 |  |
| **Moderated by:** |  |  |  |
| **Reviewed by** |  |  |  |
| **Approved by:** |  |  |  |

**DOCUMENT LOG**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Issue*** | ***Date*** | ***Comment*** | ***Author/Partner*** |
| **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 writing 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 finalized first draft version | Eise van Maanen (FAO), Anton Ellenbroek (FAO) |
| **v.8** | 29/02/2016 | Draft version | Eise van Maanen (FAO), Anton Ellenbroek (FAO) |
|  |  |  |  |

**TERMINOLOGY**

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

**Contents**

[1 Executive summary 5](#_Toc444531697)

[2 Introduction 6](#_Toc444531698)

[3 Background 8](#_Toc444531699)

[3.1 Objectives 8](#_Toc444531700)

[3.2 Scope 9](#_Toc444531701)

[3.3 Intended Audience 9](#_Toc444531702)

[3.4 Approach 9](#_Toc444531703)

[4 Landscape/Seascape 11](#_Toc444531704)

[4.1 Legal interoperability instruments 11](#_Toc444531705)

[4.1.1 Legal interoperability 11](#_Toc444531706)

[4.1.2 Legal interoperability instruments 15](#_Toc444531707)

[4.2 Data structure and size 20](#_Toc444531708)

[4.3 Global initiatives 21](#_Toc444531709)

[4.4 EU initiatives 24](#_Toc444531710)

[5 Stakeholder Consultation 26](#_Toc444531711)

[5.1 Identification of Stakeholders 26](#_Toc444531712)

[5.2 Stakeholder classification 27](#_Toc444531713)

[5.3 Marine Data Coverage 28](#_Toc444531714)

[5.4 Data Flows 29](#_Toc444531715)

[5.4.1 Data usage and citation 29](#_Toc444531716)

[5.4.2 Data management 30](#_Toc444531717)

[5.4.3 Infrastructure support to data flows 30](#_Toc444531718)

[5.5 Stakeholder Findings Representing Opportunities for EGI 31](#_Toc444531719)

[6 Interview Analysis 33](#_Toc444531720)

[7 BlueBRIDGE Case 37](#_Toc444531721)

[8 Findings and Opportunities 38](#_Toc444531722)

[9 Conclusions 40](#_Toc444531723)

[10 Bibliography 41](#_Toc444531724)

# 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 NA2.7, i.e. the 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. This document in particular provides a clear analysis of the legal barriers of data sharing within the fishery and marine sciences sector; and legal interoperability instruments to take away these barriers.

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.

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 interoperability 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 13 interviews with people working for data sharing institutes such as iMarine, IRD, ICES, Ecopath and FAO. The primary objective of the interviews was to gather information on the current legal barriers to share data in the fishery and marine sciences datasets (raw and processed data) with other institutes and communities. The interviews reveal the following legal barriers to share data and recommendations: lack of understanding of rules to share data; lack of resources; being first to publish appears to be a major incentive to withhold data; citation should be standardized through a European citation guide; need for clear rules through a top-down decision; public institutions that funded data collection should keep copyright; and data sharing is all about trust, more attention should be given to normative standards.

BlueBRIDGE was selected as best-practice case [elaborate once chapter 7 is finished]

# Introduction

The United Nations Universal Declaration of Human Rights posited already in 1948 the access to information as a fundamental human right. 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 Market report on the Fishery and Marine Sciences Data Analysis Sector) 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 harmonization 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 capabilities[[1]](#footnote-1).

This document works towards this aim through WP2 Strategy, Policy and Communications under the task NA2.6. 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 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 characterizes the stakeholders. The focus of this stakeholder analysis is on the fishery sector. It also contains high level recommendations.

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.

# Background

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

## Objectives

This study is part of the second work package of EGI-Engage (EGI-Engage:WP2 (NA2) Strategy, Policy and Communications) with as 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 Market report on the Fishery and Marine Sciences Data Analysis Sector (written by Nadia Nardi).

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.

## Scope

This study will give an introduction in data sharing policies and legal frameworks within the fishery and marine sciences sector. We assume that the fishery and marine sciences sector includes fisheries[[2]](#footnote-2), aquaculture[[3]](#footnote-3) and maritime sectors.

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. At first, the focus will be on legal interoperability of research data in similar sectors, and subsequently the scope will be reduced to 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 inevitable.

## 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.

## 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 successfully serves the fishery and marine sciences data analysis domains. A senior expert has also been involved in writing of this document.

# Landscape/Seascape

There are many benefits to be derived from open data. At first, sharing public data without any restrictions has proven to be economic, creating value multiple times and providing much greater returns on the public investment than restrictive approaches (GEO, 2015). Furthermore, open data is social welfare maximizing, and has several side-effects such as diverse reputational benefits for data owners. Placing data in the public domain also facilitates 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. 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.

## Legal interoperability instruments

### 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’*[[4]](#footnote-4)*.* 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 flavours 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. It is expected that the stimulation of legal interoperability also fosters technical 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.*

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 data[[5]](#footnote-5) 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 harmonized 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
	1. Copyright
	2. Database protection right
	3. 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, an equilibrium must be balanced out between the public interest on full legal interoperability and thus free information exchange, against the conflicting legitimate interests. 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 an original work exclusive rights for its use and distribution for a limited amount of time. 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 have even less or no eligibility for copyright protection.

Unfortunately there is no international version 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 high[[6]](#footnote-6).

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 nationals[[7]](#footnote-7) [discuss validity of copyright on metadata].

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.

*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 recognized 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 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 which 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-utilization 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). 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 jeopardize 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 which 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 characterized 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 to share data.

### Legal interoperability instruments

A variety of laws and regulations restrict the access, use and reuse[[8]](#footnote-8) 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 authorization 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 specialized 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 harmonization); at the other hand, there is a need for legal interoperability between data providers (such as licensing).

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

#### International legal interoperability instruments

**Standardization**

Standardization 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 recognized 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. Standardization can be seen as a first step to harmonization.

**Harmonization**

Harmonization is a step further in the direction of legal convergence than standardization. Regulatory harmonization can be defined as the legal model for institutionalizing 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).

Harmonization 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.

**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 harmonization, but is challenging to implement with many countries involved.

#### 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**

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. Licenses 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).

Data licensing rules are only enforceable vis-a-vis other contractual parties, as opposed to data ownership which is enforceable against third parties. In reality however, data users accept many licensing agreements by merely clicking or even by accessing a web page via internet.

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 organization 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. 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 licensor[[9]](#footnote-9). 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 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 |
| Creative Commons Zero | Anyone | Yes, with no restrictions whatsoever. | Ideal. |

*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 terms of use agreement is legally binding and may be subject to change. A terms 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.

Aggregated data, embargos and other Access Control Measures

Life cycle management and Authorization and Quality control checks

Contributions expected from [ Paul Uhlir] and [FAO]

## 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.

[state differences between categories in complexity of legal interoperability etc]

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 copyright[[10]](#footnote-10). 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 organize 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 characterized 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 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.

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, require big data collection. This will form another practical challenge to share data.

Contributions from [Julien Barde], [Jeroen Steenbeek], [Lino Pasquale Pagani] and [Marc Taconet]

## Global initiatives

There are many initiatives encouraging and supporting institutions to share data. This analysis focuses on the fishery and marine sciences sector. But other global initiatives relevant for their approach to legal interoperability are also discussed.

**iMarine**

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.

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 jeopardizing 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 BiolCube (generate new knowledge from data), the ConnectCube (collaborative, standards-oriented data publication environment, including semantic technologies), GeosCube (collaborative, standards-oriented data publication environment, including semantic technologies) and the StatsCube (data-life-cyclesupporting 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.

**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.

**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 domain[[11]](#footnote-11).

Major data publishers for GBIF 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)**

The 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. Uhlir (please see chapter six for an interview with Mr. Uhlir), 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 achieved[[12]](#footnote-12).

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.

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).

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.

## 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**

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 curves[[13]](#footnote-13).

**Infrastructure for Spatial Information in the European Community (INSPIRE)**

INSPIRE 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)**

As mentioned before, ‘big data’ will gain relevance in future years also in the fishery and marine sciences sector. The objective of EMBL-EBI 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. We play a key role in ELIXIR, the emerging infrastructure for biological information in Europe, and BioMedBridges, a project to build technical bridges between data and services in the biological, medical, translational and clinical domains. We are also pivotal partners in many other initiatives that impact the global scientific community.

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

ICES is an intergovernmental organization 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 reproduction 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 instute, 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**

Lifewatch is a European Research Infrastructure for Biodiversity that serves as an integrated access to a variety of fata, analytical and modelling tools as served by a variety of collaborating initiatives.

DG Mare? EStat? IRD? FLUX? Forth? Ecoscope? FIRMS? AgInfra?

Contributions from [Johannes Keizer] and [Marc Taconet]

# Stakeholder Consultation

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.

## Identification of Stakeholders

The domains of interest in fisheries and marine science was 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 recognize three main categories, similar to those of D2.7:

* Monitoring; national public sector organizations monitoring fisheries
* Management; (a) provide fisheries management recommendations and (b) provide plans and measures to national and regional governmental organizations
* 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 Organizations
4. Monitoring Control and Surveillance Organizations
5. International Organizations

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.

## 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 | * Activity - deliver fishing licenses, define fisheries management measures, provide recommendation for regulation
* Scope of data - national
* Using LI instruments
 |
| Fisheries Research Institute | * Activity - monitor fisheries, provide recommendations for fisheries management
* Scope of data - national data
* Using LI instruments, and using LI infrastructure services
 |
| Regional Fisheries Management Organization (RFMO) | * 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 Organization | * 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 Organization  | * Activity - centralize data through definition of data calls and standards, which are shared with state members, fishery management (harmonization) for EU level with a global impact
* Scope of data - EU and global view (fleets operating at globally)
* LI definition, using LI instruments, and using LI infrastructure services
 |

## 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.



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.

## 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 his 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.

### Data usage and citation

The legal interoperability issues related to direct exploitation of data:

* **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.

### 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 mangers implement the copyright policy of their organizations, 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 organizations 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.

### 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 organization, which may complicate the legal framework defining ownership.
* **Data location**; Most owners currently refer to a custodian within the (often literally) walls of an organization. Reformulating an organization 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 a reconsideration at organisational and collaborative level on how best to leverage the advantages of infrastructure based services to meet new requirements for legal interoperability.

## 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 summarized 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, harmonize, 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 ‘harmonized’ 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 specialized 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 standardizing fisheries reporting formats from all stakeholders (Member States, RFMO, etc.) through UN/CEFACT[[14]](#footnote-14) standards and this could create business

# 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 |
| Julien Barde | Researcher | IRD |
| Aureliano Gentile | Information Manager | FAO-FIPS |
| Nadia Nardi | Researcher and Project Manager | ENG |
| Johannes Keizer | Information Systems Officer | FAO-OPCC |
| Jeroen Steenbeek | Software Engineer | Ecopath |
| Laura Pasetto | Legal Officer | FAO-LEG |
| Blaise Kuamlangan | Chief | FAO-LEG |
| Paul Uhlir | consultant in research data policy and management | N/A |
| Marc Taconet | Chief | FAO & iMarine Board Chair |
| Neil Holdsworth | Head of Data Centre | ICES & iMarine Board member |
| Nathalie Morcrette -S | Head of Intellectual Property and Contracts Unit | INRA Science & Impact |
| Andres Vatter Rubio -S | Legal Officer | FAO-LEGN |
| Enrique Alonso García-S | Counsellor of State Head of the Division of Innovation & Science and Environmental, Rural & Marine Affaires | Consejo de Estado |

*S=scheduled interview, not yet processed in this first draft version*

**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 misuse for derivate products is an important reason for the reluctance to share data.

Pasquale Pagano (Researcher, ISTI Institute) brings about 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 Uhlir (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 Uhlir.

Uhlir also discusses the lack of knowledge and capacity to understand legal agreements, which is conditional for data sharing. Uhlir 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. Uhlir 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.

**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 standardizing 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 standardized versions of open data.

Neil Holdsworth (Head of the Data Centre, ICES) states 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, 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.

**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 standardized through a European citation Guide**

Uhlir, Keizer, Holdsworth and Steenbeek state that a standardized citation guide would stimulate data sharing and thus legal interoperability.

**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 states 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.

**Public institutions that funded data collection should keep copyright**

Uhlir recommends to develop a top-down default rule of sharing data produced by public institutions. Barde states 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 wants 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.

**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 categorization 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.

# BlueBRIDGE Case



Use case/best practice example

How does this case function in reality?

Contributions expected from [Lino Pasqaule Pagani] and [Jeroen Steenbeek]

# Findings and Opportunities

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.

1. Lack of resources major reasons for reluctance to share data
2. Being first to publish results appears to be a major incentive to withhold data
3. Citation should be standardized through a European citation Guide
4. There is a need for clear rules through a top-down decision
5. Public institutions that funded data collection should keep copyright
6. Data sharing is all about trust, more attention should be given to normative standards
7. As long as relevant government measures are not adopted and enforeced 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.
8. Provide free open data to data users in data-poor areas such as developing countries where they cannot afford the data.

[mention examples]

1. There should be a quality certificate for data

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.

|  |  |  |
| --- | --- | --- |
| Problem | Action | Responsible |
| Lack of understanding of rules to share data |  |  |
| Lack of resources major reasons for reluctance to share data |  |  |
| Being first to publish results appears to be a major incentive to withhold data |  |  |
| Citation should be standardized through a European citation Guide |  |  |
| There is a need for clear rules through a top-down decision |  |  |
| Public institutions that funded data collection should keep copyright |  |  |
| Data sharing is all about trust, more attention should be given to normative standards |  |  |
|  | Provide free open data to data users in data-poor areas such as developing countries where they cannot afford the data. |  |
| There should be a quality certificate for data |  |  |

Relevance and applicability of legal interoperability to data driven developments

Obstacles and potential in relevant markets / scenarios / studies

Recommendations for EGI life-cycle management of legally interoperable data-flows

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sector | Technology/Policy Area | Recommendation/Requirement | Priority | Timeline |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

#  Conclusions

#  Bibliography

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

*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 Costal 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, Ariadnse 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

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*

1. https://www.egi.eu/about/egi-engage/ [↑](#footnote-ref-1)
2. 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. [↑](#footnote-ref-2)
3. 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. [↑](#footnote-ref-3)
4. <http://dictionary.cambridge.org/dictionary/english/interoperability> [↑](#footnote-ref-4)
5. 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 [↑](#footnote-ref-5)
6. 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. [↑](#footnote-ref-6)
7. <http://www.wipo.int/treaties/en/text.jsp?file_id=283698> [↑](#footnote-ref-7)
8. 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.” [↑](#footnote-ref-8)
9. https://creativecommons.org/Version4/ [↑](#footnote-ref-9)
10. http://data.research.cornell.edu/content/intellectual-property [↑](#footnote-ref-10)
11. <http://www.canadensys.net/2012/why-we-should-publish-our-data-under-cc0> [↑](#footnote-ref-11)
12. https://rd-alliance.org/about.html [↑](#footnote-ref-12)
13. https://www.egi.eu/about/egi-engage/overview.html [↑](#footnote-ref-13)
14. http://www.unece.org/cefact [↑](#footnote-ref-14)