iMarine Community best practices - DRAFT

Data access and sharing policy

Revision history		
Date	**Author**	**Summary**
19 March 2013	M. Taconet, A. Gentile, A. Ellenbroek	First creation of the document.
October 2013	M. Taconet, A. Gentile, A. Ellenbroek	Addition of the preamble and draft best practices arranged by topics.
March 2014	P. Pagano, J. Ritter, A. Gentile, A. Ellenbroek	Addition of comments and new sections including meta data formats.
June 2014	M. Taconet, A. Gentile, A. Ellenbroek	Consolidated version, addition of new sections including data usage metrics, preservation policy, infringement notification and links to guidelines.
September 2014	M. Taconet, A. Gentile, A. Ellenbroek	Final version of the document produced for the closure of the project.

1. Purpose of this document

This document complements the chapters "terms of use" and "privacy policy", of the iMarine Data access and sharing policy.

This document is at a draft stage of collecting policy principles and best practices for the proper utilization of the iMarine data e-infrastructure along its terms of use. It has been compiled by harvesting relevant feedback from the Board 2, 3, 4 & 5 meeting reports. The introductory section lists best practice elements which have been identified by the Board. Those best practices items are at different stages of development.

1.1 Revision History

September 2014, the fourth and final version of the document is presented for approval by the iMarine Board members before the end of the project. It also includes a new section on exploitation agreements, links to guidelines in new Wiki pages or paragraphs.

June 2014, a third version of the document is presented to the iMarine Board members; it also includes new sections (e.g. data usage metrics, preservation policy, infringement notification and links to guidelines).

October 2013, a second version of the document is presented to the iMarine Board members.

March 2013, a first version of the document is presented to the iMarine Board members.

2. Community best practices

Table of Contents

1. Introduction

The iMarine "Community best practices" complements the iMarine Policy (i.e. "Terms of Use" and "Privacy Policy"), with a focus on the best utilization of the iMarine Data e-Infrastructure in order to fulfill the goals of the iMarine community. These goals rely on a set of principles:

- **1.1 Commitment to Open data strategy**, with a focus on recognition of data owners' full control and sovereignty over their information contribution, and credit to the intellectual ownership;
    - Mandatory terms and conditions are set forth in the Terms of Use, e.g. for ensuring that published products do not infringe any of policy of data originators, or for ensuring citation for full acknowledgment and credit to authors

  Best practices include:

    - the scoping of the uploaded, shared or published data which has to comply with the iMarine objectives
    - an iMarine Business Metadata policy
    - provisions for confidentiality and security requirements in order to encourage the participation of the industry in contributing its knowledge to the scientific process

iMarine Community best practices

- **1.2 Commitment to Open Access strategy**, with a focus on facilitating users' experience in accessing as comprehensive as possible sources of "Ecosystem Approach to Fisheries Management and Conservation of Marine Living Resources" (EA) data, promoting the principle of timeliness of the data made available, and delivering performing data-access-response-time services.

  Best practices include:

  - a Metadata formats policy
  - a re-distribution policy

- Information contributors will individually benefit from contributing data to iMarine;

  Best practices cover:

  - a Data usage metrics policy
  - a SLA policy
  - a high level of usability policy (e.g. data collection policy)

- The EA-Community as a whole will benefit from contributions of its members.

  Best practices cover:

  - content quality and timeliness
  - derivative products
  - data preservation policy
  - the acknowledgment of the iMarine Data e-Infrastructure

Best practices presented in this document are further extended with related documented guidelines. Pointers to these guidelines are provided when available.

2. Exploitation agreements

Exploitation agreements can be defined between selected users and iMarine to complement the default policies. Memorandum of Understanding - MoU, Service Level Agreement - SLA or a simple letter of agreement can be compiled to cover particular cases of exploitation of the iMarine Data e-Infrastructure.
Examples of SLAs are available in the iMarine Wiki at http://wiki.i-marine.eu/index.php/Exploitation_agreements.

----

3. Metadata

3.1 Business metadata policy

Under the Open Data principle, iMarine:

- recognizes the primary responsibility to data owner to define the applicable policy;

- promotes availability of data under existing data policies, and with proper documentation to let users understand the meaning of the accessed data, and

- promotes Citation and due provision of credit.

iMarine's recognition of these Open data principles requires a policy of Business metadata. Business metadata consists of a minimum set of descriptive metadata implementing the Open data principles and accompanying each iMarine shared dataset and information resource, if applicable. The following is an exhaustive list of the information attached to any shared content available in the iMarine Data e-Infrastructure.

Ownership and context

- Owner
- Context

Authorship

- Author
- Title
- Publisher
- Creation date
- Last update date
- Expiry date
- Contact

Copyright licenses

- Rights management, Creative Commons License type (or other licenses …)

Content description

- Description
- Spatial Scale
- Coverage
- Language
- Custom bibliographic citation
- Media type
- Identifier

The Business metadata policy incorporates various best practices by:

- requiring metadata when the dataset is shared;
- Making metadata applicable at dataset level (i.e. a larger corpora of data), with inheritance at records level
- Aiming at complete business metadata. Considering the frequent problem of absence of the required Metadata at data providers' level, iMarine encourages data providers to consider their license and terms of use policies (clearly stating these as the responsibility of the data providers) and to map their Terms of use with CC; iMarine could consider facilitating providers task in implementing CC in their websites through making available semantic technologies applied to the CC framework.

Proposed implementation of business metadata and related guidelines are described in the iMarine Wiki at http://wiki.i-marine.eu/index.php/Metadata_standards#Dublin_Core_and_Business_metadata.

3.2 Metadata - formats and standards policy

iMarine's application of the Open Access concept focuses on facilitating users experience in accessing as comprehensive as possible sources of EA data, promoting the principle of distributed data infrastructure together with that of timeliness of the data made available, and delivering performing data-access-response-time services. These principles require the implementation of a set of interoperability and transformation services building on Metadata formats and standard. Since implementing services on a given Metadata format is costly and is likely to require prioritization, the policy covers the following aspects:

- Criteria for eligibility of a Metadata format (e.g. standard widely used among EA CoP, a format indispensable for optimized use of the iMarine Data e-Infrastructure)
- Criteria prevailing for levels of implementation of support tool for metadata format. Discussion on aspects related to those criteria and categorization can be found at http://wiki.i-marine.eu/index.php/Metadata_standards
- Categorization and documentation of available services for each Metadata format and standard

Metadata currently supported:

- Biological and ecological data are made available in Darwin Core. DwC is a de-facto standard to represent data and metadata;

- Taxonomic information are made available in Darwin Core Archive format. DwCA is a de-facto standard to deliver archives of data expressed in DwC;

iMarine Community best practices

- Geo-referenced data are described using ISO 19115/19119 and made available through the OGC protocols: WMS, WCS, WFS, etc. Biological and ecological data can be also enriched with ISO 19115/19119 if they include a geographic coverage;

- Statistical Data are made available in SDMX that includes specific metadata for agencies, code lists, and datasets;

- Textual documents, such as articles, journals, papers, etc. are described using Dublin Core.

The on-going discussion on Metadata formats is available in the iMarine Wiki at http://wiki.i-marine.eu/index.php/Metadata_standards.

4. Best use of Creative Common (CC) licenses

Beyond the original license scheme, a type of CC license must be assigned anytime content is exposed among VREs or any time a product is made available publicly. The license options are described at http://creativecommons.org/licenses/.

For products intended for public use, where feasible, the adoption of the CC-By license is recommended, consistent with the openness of data as an iMarine best practice. As per Terms of Use, users undertake to comply with their institutions' policies and copyrights including versions of open license adopted by their institutions.

5. Option for exposing content: Sharing and Publishing

Content is public when exposed to general view with no restrictions (except for possible registration procedures), when openly accessible. When users in the iMarine Data e-Infrastructure publish a content, it means that the content is exposed to the general public (e.g. through a web service). Exposing content within the iMarine Data e-Infrastructure (i.e. sharing through one or more VREs) is a different action in which the content is only available to a limited audience under restricted access.

Access to a given VRE is granted to users by the VRE administrator, according to the related VRE policy, which shall be advertised at VRE entry level. The best practice for VREs of general purpose is to grant access to users only if its content is relevant to the activity and tasks of the applicant.

iMarine users can share and publish data, but are in any case expected to commit to the data owner's sharing and publishing policy, and to the attached CC license and VRE policy.

iMarine Community best practices

----

6. Derivative product - Secondary use

A derivative product can be the results of a:

- ○ Mashing up process where data from different sources are put together and connected with each other (e.g. AppliFish);
- ○ Computing process utilizing data from different sources to produce new analysis and results (e.g. SPREAD statistical reallocation).

Figure 1 shows licenses for derivative products according to data manipulation.

When a derivative product is produced through mashing up among two or more owners, it is a good practice to identify the "lead creator" that will be the contact point for any requests related to the new product. The lead creator should be a legal body that is the legitimate **copyright holder**, owns the derivative product and is contact point for any use, re-use/re-distribute request. By default, the VRE owner will be considered the lead creator.
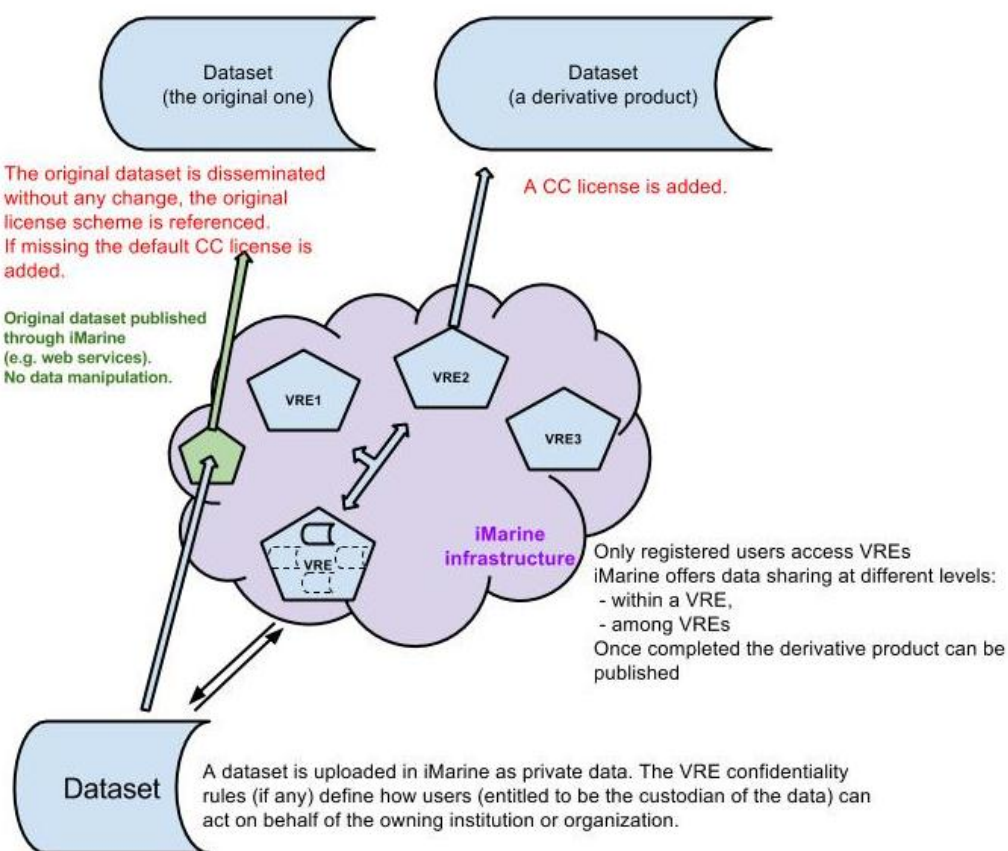

Figure 1 - Data manipulation and licenses for derivative products.

7. Data collection policy

A Data collection is a set of data under an ownership and following consistent rules such as homogeneity, quality standards, production process and workflow, dissemination format, etc.

Each data collection accessible through iMarine is documented with its associated policy, called "data collection policy", which enables users to evaluate the information content and conditions for utilization. Data owners are providers and responsible for the maintenance of such documentation.

A data collection policy should include mandatory and optional fields.

- Mandatory: business meta data and key quality criteria (e.g. sources, timeliness).
- Optional: any other element enabling users to evaluate the quality of the dataset.

All elements of a data collection inherit the properties of the data collection and can be cited. The citation includes reference to the data collection policy.

Example

In the FAO Fisheries and Aquaculture Department (FI) and in the Fishery Resources Monitoring System (FIRMS) similar documentations are called "Data collection" and "Data quality assurance".

See in example:

- FI Data collection fact sheets http://www.fao.org/fishery/statistics/collections/en
- FIRMS Data quality assurance fact sheets http://firms.fao.org/firms/contributions/en

8. Citation policy

iMarine information and data products are disseminated with a "how to cite" feature. All published content should be accompanied by its preferred citation. For this purpose, the following fields should be specified:

- Corporate Author/ Personal Author
- Reference year
- Title (cover page and chapter titles)
- Website identity
- Physical site location
- Date of update
- Web page link

iMarine Community best practices

> **Example**
>
> - Citing a fact sheet in AppliFish:
>
> © FAO, 2013. AppliFish. Albacore. In: iMarine (Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources) [online]. Updated 12 march 2013. [Cited 14 March 2013]. http://www.i-marine.eu/AppliFish/
> See "Disclaimer and Copyright Notice" for the whole list of contributors to AppliFish.
>
> - Citing a website:
>
> International Maritime Organization (IMO). 2011. IMO [online]. London. [Cited 28 June 2011]. www.imo.org
>
> The application will add at the end of any citation the following statement: "*iMarine cannot vouch for the data or analyses derived from these data after the data have been retrieved from i-Marine.eu*."

In case of citation for derivative products it is best practice to mention the originators of the data with which the derivative products have been realized. This is in fact required for any type of CC license, except for the CC0 license. Based upon the logic that any shared or public content in the iMarine Data e-Infrastructure goes with its own citation, the citation mechanism is foreseen to provide ways to concatenate all citations for the utilized sources (dataset, documents etc.).

See also a draft template for a citation of a derivative product in the iMarine Wiki at http://wiki.i-marine.eu/index.php/Content_citation#Template_for_a_citation_of_a_derivative_product

9. Acknowledgment to iMarine

Content elaborated through iMarine is accompanied with a "Powered by" statement acknowledging the role of iMarine platform in its generation. It is good practice to acknowledge iMarine as appropriate (e.g. "Credits") for the benefits and visibility of the iMarine community itself.

10.   Data usage metrics

In order to properly inform contributors on usage of their product, as well as for accounting and auditing purposes, content exposed to the public through the iMarine Data e-Infrastructure is monitored for usage metrics measuring use and distribution of Apps, visits to VRE.

The description of the usage metrics computed by the Infrastructure is available at http://wiki.i-marine.eu/index.php/Procedure_Infrastructure_Accounting.

iMarine Community best practices

----

11.   Preservation policy

Protecting ownership of data and preserving access to such data is a major concern for the iMarine Consortium. If any service will be discontinued, where reasonably possible, users will receive reasonable advance notice and a chance to get information out of that Service. For this reason it is very important to keep users' contact information up to date.

iMarine commits for long term maintenance of the content made available through iMarine's supported Metadata formats.

12.   Submit an infringement notification

If you believe your copyright-protected work was misused within the iMarine Data e-Infrastructure, you may submit an infringement notification to info@i-marine.eu. These requests should only be submitted by the content owner or an agent authorized to act on the owner's behalf.
Your submission must contain your contact information, a description of your work that you believe has been infringed including the URL where such content is available in the Infrastructure ("Get Link" from the Workspace), and your physical or electronic signature.
Your submission will contain also the statement: "I have a good faith belief that the use of the material in the manner complained of is not authorized by the copyright owner, its agent, or the law." And the following statement: "The information in this notification is accurate, and under penalty of perjury, I am the owner, or an agent authorized to act on behalf of the owner, of an exclusive right that is allegedly infringed."

13.   Glossary

- **iMarine Consortium (the Consortium)**
  iMarine project is co-funded by the European Commission under Framework Programme (FP7). The project was launched in November 2011 and will end in August 2014.
  Aims to establish and operate an e-infrastructure supporting the principles of the ecosystem approach to fisheries management and conservation of marine living resources.
  Partners currently part of the iMarine Consortium are:
    - Centro de Referencia em Informacao Ambiental (CRIA), Brasil
    - Consiglio Nazionale delle Ricerche (CNR), Italy
    - Engineering Ingegneria Informatica SpA (E-IIS), Italy
    - European Organisation for Nuclear Research (CERN), Switzerland
    - Fishbase Information & Research Group Inc. (FIN), Philippines
    - Food and Agriculture Organisation of the United Nations (FAO), Italy

- -

○ Foundation for Research and Technology Hellas (FORTH), Greece
○ GEIE ERCIM (ERCIM), France
○ Institut de Recherche pour le Developpement (IRD), France
○ National and Kapodistrian University of Athens (NKUA), Greece
○ OBIS, United Nations Educational, Scientific and Cultural Organization – UNESCO (UNESCO), France
○ Terradue s.r.l. (Terradue), Italy
○ Trust-IT Services Ltd. (Trust-IT), United Kingdom

- **iMarine Data e-Infrastructure (the Infrastructure)**
  iMarine utilizes the distributed infrastructure of D4Science (D4S - DIstributed colLaboratories Infrastructure on Grid ENabled Technology 4 Science) to provide computational resources to a target community of users. iMarine resources include hardware, datasets, and software services. The datasets are the input and output of those processes, hence the core digital assets of the community. The services capture the processes that are directly or indirectly relevant to the needs of the target community. Hardware resources is where processes unfold and datasets are stored. The enabling technology for D4Science is gCube.

- **iMarine Gateway**
  The access point to iMarine products and services.

- **Virtual Research Environment (VRE)**
  In iMarine, collaboration services and related access to content are available through Virtual Research Environments (VREs). These VREs can take various forms including web interactive user interface, web applications, and pluggable standalone user interface.
  VRE, as a collaborative environment, allows scientists and practitioners to access cross-disciplinary dataset and knowledge, as well as a rich array of services that enable innovative analysis, visualization and domain specific knowledge generation processes. VREs can be created for a specific period and a specific task and only authorized users access data and services exposed through these VREs.

- **User**
  Individuals acting on behalf of institutions or organizations that have entered into an agreement with the iMarine Consortium granting them access to the iMarine Data e-Infrastructure, or individuals who were otherwise granted access by the iMarine Consortium. Any agent (human or software) granted to access iMarine content is a user. Data providers and data consumers are here considered as users. From the infrastructure perspective, users own all data they upload.
  The user means the individual (or entity) that creates the Content.

- **Content**
  The totality of information and data in any digital format contained in the iMarine web

iMarine Community best practices

----

infrastructure.

- **Derivative Work**
  Whatever content produced in whatever format based upon one or more resources taken from the iMarine Data e-Infrastructure.

- **Business Metadata**
  A minimum set of descriptive metadata to identify uniquely the content and to know the associated license scheme.

- **Public Content and Shared Content**
  Content is **public** when exposed to general view with no restrictions (except for possible registration procedures), when openly accessible. When users in iMarine publish a content means that it is exposed to the general public (e.g. through a web service). **Sharing** content is exposing content within the iMarine Data e-Infrastructure (i.e. sharing through one or more VREs), and is a different action in which the content is only available to a limited audience under restricted access.

- **Workspace Service**
  A virtual space that allows users to gather, organize and exchange files across VREs.

- **Email Service**
  A web service enabling messages exchange only among iMarine registered users.

  Social Service
  A web service to share information by allowing users to post updates, including links to news articles and new derivative datasets, and to comment on content published by some of the iMarine Applications or Users.

- **Notification Service**
  For the purpose of facilitating the user interaction with the iMarine Gateway Social Service and Email Service, a Notification Service is also present. The Notification Service can be configured from iMarine Gateway to an email address associated with your account

14.   Links to guidelines

Guidelines aim to constitute the hands-on manual to allow users to best exploit the iMarine Data e-Infrastructure. Many best practices are at vision stage and related guidelines at draft stage and are not supported by the infrastructure yet.

Guidelines at maturity level

- iMarine Services: http://www.i-marine.eu/Content/OurServices.aspx?menu=1

- Metadata: http://wiki.i-marine.eu/index.php/Metadata_standards

iMarine Community best practices

- OGC/ISO Publishing guidelines for Data and Services Providers http://wiki.i-marine.eu/index.php/OGC/ISO_Publishing_guidelines_for_Data_and_Services_Providers

- Citation for a derivative product http://wiki.i-marine.eu/index.php/Use_Cases_for_EA-CoP_Data_Access_and_Sharing_Policies#Applifish_-_FAO_Species_fact_sheet

Guidelines at early stage

- Code lists management http://wiki.i-marine.eu/index.php/Use_Cases_for_EA-CoP_Data_Access_and_Sharing_Policies#Guidelines

- Content citation including citation for derivative products http://wiki.i-marine.eu/index.php/Content_citation

- [work in progress - others to be added]