



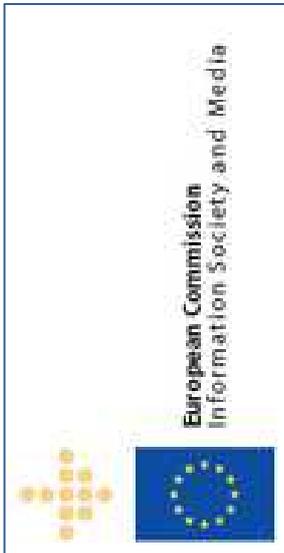
<i>Project Acronym</i>	<i>iMarine</i>
<i>Project Title</i>	<i>Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources</i>
<i>Project Number</i>	<i>283644</i>
<i>Deliverable Title</i>	<i>Data Management Software</i>
<i>Deliverable No.</i>	<i>D9.3</i>
<i>Delivery Date</i>	<i>01 2014</i>
<i>Author</i>	<i>Gianpaolo Coro (CNR), Lucio Lelii (CNR), Andrea Manzi (CERN), Nikolaos Drakopoulos (CERN), Valentina Marioli (CNR), Roberto Cirillo (CNR), Fabio Simeoni (FAO), Alexandros Antoniadis (NKUA), Fabrice Brito (Terradue), Herve Caumont (Terradue)</i>

Abstract: *The iMarine Data Management Software comprises a number of components and subsystems offering facilities for accessing, transferring and harmonising a rich array of data typologies. This document briefly describes the novelties included in the iMarine Data Management Software from M12 (Oct. '12) to M27 (Jan. '14) and offers a series of links to the accompanying documentation.*

DOCUMENT INFORMATION

PROJECT	
Project Acronym	iMarine
Project Title	Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources
Project Start	1st November 2011
Project Duration	30 months
Funding	FP7-INFRASTRUCTURES-2011-2
Grant Agreement No.	283644
DOCUMENT	
Deliverable No.	D9.3
Deliverable Title	iMarine Data Management Software
Contractual Delivery Date	01 2014
Actual Delivery Date	01 2014
Author(s)	Gianpaolo Coro (CNR), Lucio Lelii (CNR), Andrea Manzi (CERN), Nikolaos Drakopoulos (CERN), Michal Simon (CERN), Valentina Marioli (CNR), Roberto Cirillo (CNR), Fabio Simeoni (FAO), Alexandros Antoniadis (NKUA), Fabrice Brito (Terradue), Herve Caumont (Terradue)
Editor(s)	Gianpaolo Coro (CNR)
Reviewer(s)	L. Candela (CNR), P. Pagano (CNR)
Contributor(s)	N/A
Work Package No.	WP 9
Work Package Title	iMarine Data Management Facilities Development
Work Package Leader	Gianpaolo Coro (CNR)
Work Package Participants	CNR, CERN, NKUA, FAO, Terradue
Estimated Person Months	23.40
Distribution	Public
Nature	Other
Version / Revision	1.0
Draft / Final	Final
Total No. Pages (including cover)	9
Keywords	Data Transfer, Data Storage, Data Harmonization, Data Access

DISCLAIMER



iMarine (RI – 283644) is a Research Infrastructures Combination of Collaborative Project and Coordination and Support Action (CP-CSA) co-funded by the European Commission under the Capacities Programme, Framework Programme Seven (FP7).

The goal of iMarine, *Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources*, is to establish and operate a data infrastructure supporting the principles of the Ecosystem Approach to Fisheries Management and Conservation of Marine Living Resources and to facilitate the emergence of a unified Ecosystem Approach Community of Practice (EA-CoP).

This document contains information on iMarine core activities, findings and outcomes and it may also contain contributions from distinguished experts who contribute as iMarine Board members. Any reference to content in this document should clearly indicate the authors, source, organisation and publication date.

The document has been produced with the funding of the European Commission. The content of this publication is the sole responsibility of the iMarine Consortium and its experts, and it cannot be considered to reflect the views of the European Commission. The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

The European Union (EU) was established in accordance with the Treaty on the European Union (Maastricht). There are currently 27 member states of the European Union. It is based on the European Communities and the member states' cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice, and the Court of Auditors (<http://europa.eu.int/>).

Copyright © The iMarine Consortium 2011. See <http://www.i-marine.eu/Content/About.aspx?id=6cc695f5-cc75-4597-b9f1-6ebea7259105> for details on the copyright holders.

For more information on the project, its partners and contributors please see <http://www.i-marine.eu/>. You are permitted to copy and distribute verbatim copies of this document containing this copyright notice, but modifying this document is not allowed. You are permitted to copy this document in whole or in part into other documents if you attach the following reference to the copied elements: "Copyright © The iMarine Consortium 2011."

The information contained in this document represents the views of the iMarine Consortium as of the date they are published. The iMarine Consortium does not guarantee that any information contained herein is error-free, or up to date. THE IMARINE CONSORTIUM MAKES NO WARRANTIES, EXPRESS, IMPLIED, OR STATUTORY, BY PUBLISHING THIS DOCUMENT.

GLOSSARY

ABBREVIATION	DEFINITION
iMarine	Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources
WCS	Web Coverage Service
WMS	Web Map Service
WFS	Web Feature Service
SOAP	Simple Object Access Protocol
HTTP	HyperText Transfer Protocol
FTP	File Transfer Protocol
GHN	gCube Hosting Node
SDMX	Statistical Data and Metadata Exchange
DARWIN CORE	A standard designed to facilitate the exchange of information about the geographic occurrence of species and the existence of specimens in collections
ETL	Extract Transform and Load

DELIVERABLE SUMMARY

1. INTRODUCTION

The iMarine Data Management Software comprises a number of components and subsystems offering facilities for accessing, transferring and harmonising a rich array of data typologies. This document describes the novelties within the iMarine Data Management Software from M12 (Oct.'12) to M27 (Jan.'14). It complements D9.2 [10] which describes iMarine Data Management Software up to M11 (Sept.'12). This deliverable is intended for documentation purposes only. The actual deliverable is represented by the software artifacts and their accompanying documentation.

2. TARGET RELEASE(S)

The iMarine Data Management Software described by this document and thus realizing D9.3 contribute to the following releases:

- gCube 2.11.0
- gCube 2.11.1
- gCube 2.12.0
- gCube 2.13.0
- gCube 2.14.0
- gCube 2.15.0
- gCube 2.16.0
- gCube 2.17.0
- gCube 2.17.1

3. OBJECTIVES

The new versions of the components belonging to the Data Management Software, distributed as part of the referred releases, cover the following objectives:

- **Data Access Facilities:**

The activities first concentrated on enhancements in the Tree Manager Library. The main aim was to make this library compliant with the new standard communication policies of the gCube services, imposed by the adoption of the Featherweight Stack. In addition, the activities involved the implementation of a service (Species Products Discovery or SPD) able to unify information coming from several repositories about marine species, e.g. OBIS and GBIF. SPD is an infrastructural service, which is currently used by the iMarine community of practice to search among heterogeneous biodiversity datasets and to retrieve one single merged dataset, having a coherent structure. SPD is endowed with a SQL-like query language (SPQL) and with unfolding operations to enrich datasets by means of information coming from other datasets, for example synonyms and accepted names. Activities also regarded the Storage Manager facility. In particular this distributed storage system was enhanced with tools to copy, move and better manage the remotely hosted files.

- **Data Transfer Facilities:**

The activities focused on the implementation of a complete Data Transfer suite, supporting scheduling and post-processing of transferred data. The facility required enhancements in the data exchange libraries (gRS2), in order to manage several protocols like TCP, HTTP etc.

The Tree Manager system developed during the Task T9.1 activities was successfully connected and managed. The software development activity also involved a newly developed Storage Management portlet, which allows users to inspect the content of the iMarine distributed storage facility.

Data Transfer also involved the implementation of a framework for geo-spatial data processing and transfer, that is able to process geographical datasets in a distributed fashion and to deliver the results in OGC standard formats. Such framework (WPS-Hadoop) relies on the WPS OGC protocol and has been used to integrate legacy applications about Tuna fishery management (Tuna Atlas), developed by the IRD iMarine partner and delivered under the R programming language. Finally, the activities about geo-spatial data transferring focused on searching a suitable OGC description for environmental features datasets. The Sensor Observation Service (SOS) was adopted for such scope and forced to account for 4D datasets representations.

- **Data Assessment, Harmonization and Certification Facilities:**

The activities focused first on the management of the SDMX protocol for statistical data exchange, regarding in particular time series of fisheries catch quantities. These activities involved the implementation of a SDMX registry and a SDMX Data Source, integrated with the current Time Series gCube service. Effort was spent also in the enhancement of the Time Series service itself. In particular, the Tabular Data Widget component was developed to endow other services with table exploring facilities. The Tabular Data facilities were developed with accurate design and implementation phases. These involved the implementation of several library modules and of one infrastructural service (Tabular Data Manager), which act as agents of a Data warehouse system. The whole system is able to manage the time dimension at different granularities and to distribute a time series on days, weeks, years etc. An interface for the Tabular Data Manager has been released which allows users to curate, harmonize and assess their datasets having a tabular format. SMDX management is included in this system and tools for importing and exporting SDMX data were released at library and service levels.

4. COMPONENTS

In the target releases, the following components have been updated, maintained or newly introduced:

- to support the **Data Access Facilities:**
 - asfis-spd-plugin
 - brazilianflora-spd-plugin
 - catalogueoflife-spd-plugin
 - data-access.spd-model-library
 - figis-tm-plugin
 - gbif-spd-plugin
 - irmng-spd-plugin
 - itis-spd-plugin
 - ncbi-spd-plugin
 - oai-tm-plugin
 - obis-spd-plugin

- queue-manager
 - spd-client-library
 - spd-having-engine
 - spd-model-library
 - spd-plugin-framework
 - spd-plugin-fwk
 - specieslink-spd-plugin
 - species-product-discovery-service
 - species-product-discovery-stubs
 - species-products-discovery
 - species-tm-plugin
 - spql-parser
 - streams
 - tree-manager
 - tree-manager-framework
 - tree-manager-library
 - tree-manager-service
 - tree-manager-stubs
 - tree-repository
 - tree-repository-requests
 - trees
 - tree-uri-resolver
 - worms-spd-plugin
- to support the **Data Transfer Facilities**:
 - org.gcube.data-transfer.agent-library
 - org.gcube.data-transfer.agent-service
 - org.gcube.data-transfer.agent-stubs
 - org.gcube.data-transfer.common
 - org.gcube.data-transfer.common-messaging
 - org.gcube.data-transfer.scheduler-db
 - org.gcube.data-transfer.scheduler-is-interface
 - org.gcube.data-transfer.scheduler-library
 - org.gcube.data-transfer.scheduler-portlet
 - org.gcube.data-transfer.scheduler-service
 - org.gcube.data-transfer.scheduler-stubs
 - org.gcube.data-transfer.storagemanager-portlet
 - org.gcube.data-transfer.uri-resolver
 - org.gcube.data-transfer.vsf-provider-smp
 - to support the **Data Assessment, Harmonization and Certification Facilities**:
 - org.gcube.data-analysis.cube-manager
 - org.gcube.data-analysis.cube-manager-api
 - org.gcube.data-analysis.cube-manager-data
 - org.gcube.data-analysis.cube-manager-metadata

- org.gcube.data-analysis.evaluator-sql
- org.gcube.data-analysis.operation-api
- org.gcube.data-analysis.operation-csv
- org.gcube.data-analysis.operation-sdmx
- org.gcube.data-analysis.service-client-api
- org.gcube.data-analysis.service-client-impl
- org.gcube.data-analysis.tabular-data-client-library
- org.gcube.data-analysis.tabular-data-commons
- org.gcube.data-analysis.tabular-data-service
- org.gcube.data-analysis.tabular-metadata
- org.gcube.data-analysis.tabular-model
- org.gcube.data-analysis.tabular-model-postgresql
- org.gcube.data-analysis.tabular-query
- org.gcube.data-analysis.tabular-query-parameters

5. DOCUMENTATION

A detailed specification of the services documented by this report is in a number of dedicated Wiki pages [6][7][8][11]. Moreover, technical documentation covering all the aspects of the software is available at:

- Admin’s Guide [3]
- Developer’s Guide [4]
- User’s Guide [5]

Finally, for development purpose, each component is provided with a Javadoc documentation along with a direct link to the associated section in Developer’s Guide. This artifact is available at [9].

This documentation is an integral part of the actual deliverable.

6. DOWNLOAD

This document describes a deliverable of type “other”. The actual deliverable consists of the software artifacts briefly discussed. Such artifacts are available for download via a Maven repository [1] or via the ETICS repository [2].

REFERENCES

- [1] gCube Maven Repository RELEASES:
<http://maven.research-infrastructures.eu/nexus/index.html#view-repositories;gcube-releases~browsestorage>
- [2] gCube Etics Repository RELEASES:
<https://grids16.eng.it/BuildReport/>
- [3] Administrator's Guide:
https://gcube.wiki.gcube-system.org/gcube/index.php/Administrator%27s_Guide
- [4] Developer's Guide:
https://gcube.wiki.gcube-system.org/gcube/index.php/Developer%27s_Guide
- [5] User's Guide:
https://gcube.wiki.gcube-system.org/gcube/index.php/User%27s_Guide
- [6] Milestone 37: Data Access and Storage Facilities:
https://gcube.wiki.gcube-system.org/gcube/index.php/Data_Access_and_Storage_Facilities
- [7] Milestone 38: Data Transfer Facilities Specification:
https://gcube.wiki.gcube-system.org/gcube/index.php/Data_Transfer_Facilities
- [8] Milestone 39: Data Assessment, Harmonisation, and Certification Facilities:
https://gcube.wiki.gcube-system.org/gcube/index.php/Data_Assessment,_Harmonisation,_and_Certification_Facilities
- [9] gCube Distribution Site:
https://www.gcubesystem.org/index.php?option=com_distribution&view=distribution&Itemid=23
- [10] G. Coro, F. De Faveri, L. Lelii. *iMarine Data Management Software*. iMarine D9.1 Project Deliverable. June 2012
- [11] Milestone 40: Data Management Software Consolidated Specifications:
https://gcube.wiki.gcube-system.org/gcube/index.php/Data_Management_Software_Consolidated_Specifications
- [12] G. Coro, F. De Faveri, L. Lelii, A. Manzi, N. Drakopoulos, V. Marioli, F. Simeoni, A. Antoniadis. *iMarine Data Management Software*. iMarine D9.2 Project Deliverable. September 2012