Project Acronym	***iMarine***
Project Title	***Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources***
Project Number	***283644***
Deliverable Title	***iMarine Data Infrastructure Enabling Software***
Deliverable No.	***D8.1***
Delivery Date	***04 2012***
Author	***Manuele Simi – CNR***

Abstract:
iMarine Data Infrastructure Enabling Software contains the description of software and pointers to the related documentation and artifacts of the components that comprise the e-Infrastructure Management suite delivered up to M6

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).

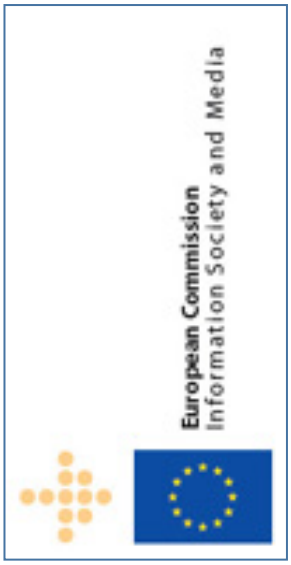
DOCUMENT INFORMATION

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.	D8.1
Deliverable Title	iMarine Data Infrastructure Enabling Software
Contractual Delivery Date	30/04/2012
Actual Delivery Date	14/05/2012
Author(s)	Manuele Simi – CNR, Gerasimos Farantatos – NKUA
Editor(s)	Manuele Simi – CNR
Reviewer(s)	Andrea Manzi – CERN
Contributor(s)	Fabio Simeoni – FAO
Work Package No.	WP 8
Work Package Title	iMarine Data e-Infrastructure Enabling-technology Development
Work Package Leader	Manuele Simi – CNR
Work Package Participants	CNR, NKUA, FAO, E-IIS
Estimated Person Months	21.70
Distribution	Public
Nature	Other
Version / Revision	1.0
Draft / Final	Draft
Total No. Pages (including cover)	10
Keywords	Enabling Software, gCube, Maven

CHANGE LOG

Reason for Change	Issue	Version	Date	Partner / Responsible
First draft		0.1	20/04/2012	CNR
Version ready for Internal Review		0.2	10/05/2012	CNR
Version ready for Official Review		0.3	14/05/2012	CNR
Version addressing Reviewer Comments		0.4	14/05/2012	CNR

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/).

GLOSSARY

ABBREVIATION	DEFINITION
iMarine	Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources

DELIVERABLE SUMMARY

1. INTRODUCTION

This deliverable describes the novelties and evolution of the iMarine Data Infrastructure Enabling Software up to M6 (Apr.'12).

2. TARGET RELEASE(S)

This deliverable reports on the software released as part of the following gCube releases:

- gCube 2.7.2
- gCube 2.8.0
- gCube 2.8.1
- gCube 2.9.0

3. OBJECTIVES

The new version of components belonging the Data Infrastructure Enabling Software released as part of the target releases covers the following objectives:

- support the gCube mavenization process

  WP8 sets out to support gCube developers that wish to use Maven to build their components. In doing so, we acknowledge the value of its mechanisms for dependency management, its infrastructure of public repositories for dependency resolution, the extensive build support offered by its plugins, and the sophistication of the tools that integrate it with popular IDEs. On the infrastructure side, WP8 is responsible for gathering access to gCube software for deployment purposes. New components have been developed to replace the previous in-house storage and access site.

- support encryption/decryption of gCube resources

  gCube resources have longed suffered of the limitation to do not held sensitive data. A new general purpose library have been now designed and developed, common-utils-encryption, to offer an easy way to encrypt and decrypt XML documents and String objects. The library has been then exploited by the existing resource management library (now part of gCore Framework) to protect the resource serialization. The library uses a symmetric key based on the AES[8] standard algorithm for cryptography.

- offer tools for *in-container* integration testing

  The notion of integration testing covers a wide spectrum beyond unit testing, from integration of service components to gCube integration over a wide-area network. Our need is particularly

stronger just one step beyond the scope of unit testing. With our stack and enabling technologies, this brings us to what people often refer to as in-container testing: client-driven acceptance tests that exercise functional features of a target service deployed within a target container. In-container testing is the first chance we get to test the integration between the individual components of the service.

- support consumers of gCube features

  Clients of gCube software now have new tools for developing consumers of features offered by gCube Services: (i) common-clients, an API for client library development, and (ii) common-gcore-clients, a specialization of the common-clients API for clients of gCube services

- improve the matchmaking and resource management capabilities of the gCube process execution engine (PE2ng)

  In the direction of supporting the evolution of workflow management components, as well as all PE2ng adaptors developed in the context of other work packages, a node selection and collocation policy library has been implemented. The library aims to provide options for additional execution policies and to support execution optimizations. The library is centered around the notions of Node Selectors and Collocation Policies. A client can use a Node Selector either to immediately select the most suitable execution node or to perform an assessment of the set of candidate nodes. Collocation Policies can function in a way as simple as to assign all tasks to the local or to a single remote node, or in a more complex way, in which case they can be used in conjunction with node selectors. In the latter case, and depending on the behavior of each policy, additional parameters which regulate the extent on which the policy can diverge from the optimal selection as defined by the underlying node selector can be used. This type of parameterization enables policies to be flexible enough to be used by applications with workflows governed by a variety of conditions while guaranteeing that they still operate in the general direction specified in their behavior.

- Improve the performance and scalability of PE2ng

  In addition to the improvements in matchmaking capabilities of PE2ng, workflow management software also includes enhancements in the direction of attaining more scalability and performance. In particular, the matchmaking strategy employed by PE2ng was improved for the execution of JDL DAG jobs in the D4Science infrastructure by exploiting the presence of common sets of requirements specified in the job description. The new strategy achieves matchmaking in constant time for the majority of jobs of this kind by minimizing the number of queries submitted to the Information System service, thereby making the system scalable with respect to submission times.

  Another major action in the direction of improving performance is the integration of PE2ng with the latest implementation of the gCube Storage Manager Service developed by WP9. Data input and output has significant impact on the performance of all adaptors which do not have the option to use streaming operations for the communication between computational nodes in the infrastructure. In these cases, PE2ng uses the storage layer to perform data staging and, depending on job descriptions, to read job input and produce the final output. The new Storage Manager

service is designed to handle high throughput operations of arbitrarily sized data objects, thereby satisfying the requirements of PE2ng, so the modification of the software layer which handles storage management in order to interface with the former translates to significant cumulative performance benefits, especially for complex execution jobs.

- Improve the execution status reporting capabilities of PE2ng

  PE2ng offers execution status capabilities based on an event mechanism. This mechanism was extended in order to include information about the successful completion of subsets of the initial execution plans which run on different worker nodes. The additional information includes the logical node name and data about the actual worker node selected, such as host name and the port which is used by PE2ng.

- Fix observed issues

  In the first six months of the project, Data Infrastructure Enabling Software was actively maintained in order to correct any issues which are encountered. One example of such an issue was the incorrect behavior of PE2ng in output storage in JDL DAG jobs involving multiple nodes, in the case of opting to store the output by publishing it to a remote endpoint, such as FTP. The issue resulted files originating from different nodes being overwritten and was successfully managed. Other examples include various fixings the Information System's publication phase on Runtime Resources and scoping rules, as well as reversing the integration previously unexploited features of the security framework from the node management functionalities.

4. COMPONENTS

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

- to support the gCube mavenization process:
  - gCore Framework 1.4.0 (now integrating fomer common-parsers 1.4.0)
  - legacy-wscore 1.4.0
  - maven-service-plugin 1.0.0
  - maven-parent 1.0.0
  - ghn-core-runtime 1.0.0
  - Software Gateway 1.0.0
  - Software Gateway client 1.0 .0
  - Resource Manager 1.1.0
  - Deployer 2.3.0
- to support consumers of gCube features (in collaboration with WP11):
  - common-clients 1.0.0
  - common-gcore-clients 1.0.0
- to support resource encryption:
  - gCore Framework 1.4.0 (now integrating former common-parsers 1.4.0)

    - common-utils-encryption 1.0.1

- to support *in-container* integration testing:

    - my-container 1.0.0

- to improve the matchmaking and resource management capabilities of PE2ng

    - MadgikCommons 1.1.0

- to improve PE2ng performance

    - WorkflowEngine 1.2.1
    - InformationSystem 1.3.0
    - EnvironmentProvider 1.2.0
    - GCubeEnvironmentProvider 1.3.0

- to improve the execution status reporting capabilities of PE2ng

    - ExecutionEngine 1.1.5
    - WorkflowEngineService 1.2.1
    - WorkflowEngineServiceClient 1.2.1

- to fix observed issues

    - ExecutionEngine 1.1.5
    - gHN Manager 1.5.0
    - ISPublisher 2.2.0
    - ISRegistry 2.1.3

5. DOCUMENTATION

A comprehensive overview of the subsystems the described components belong to is available at [6] (e-infrastructure management facilities) and at [7] (workflow management facilities).

Technical documentation covering all the aspects of the software is available at:

- Admin's Guide [2]
- Developer's Guide [4]
- User's Guide [5]

Finally, for development purpose, Javadoc documentation for each component, along with a direct link to the associated section in Developer's Guide, is available at [2].

6. DOWNLOAD

The components described in this deliverable are available for download at [1]. Direct links to each component are available at [2].

REFERENCES

[1] gCube Maven Repository RELEASES:
http://maven.research-infrastructures.eu/nexus/index.html#view-repositories;gcube-releases~browsestorage

[2] gCube Distribution Site:
https://www.gcube-system.org/index.php?option=com_distribution&view=distribution&Itemid=23

[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 33: Data e-Infrastructure Management Facilities:
https://gcube.wiki.gcube-system.org/gcube/index.php/Data_e-Infrastructure_Management_Facilities

[7] Milestone 35: Workflow Management Facilities:
https://gcube.wiki.gcube-system.org/gcube/index.php/Workflow_Management_Facilities

[8] Advanced Encryption Standard:
http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf