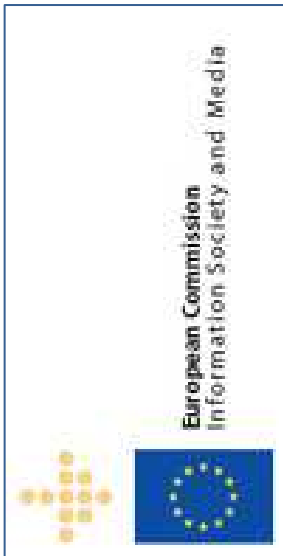| Project Acronym	*iMarine*
Project Title	*Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources*
Project Number	*283644*
Deliverable Title	*iMarine Data Infrastructure Enabling Software*
Deliverable No.	*D8.3*
Delivery Date	*10 2013*
Authors	*Alex Antoniadis (NKUA), Roberto Cirillo (CNR), Ciro Formisano (E-IIS), Luigi Fortunati (CNR), John Gerbesiotis (NKUA), Lucio Lelii (CNR), Panagiotis Liakos (NKUA), Antonio Savini (E-IIS), Fabio Simeoni (FAO), Ermannno Travaglino (E-IIS).*

Abstract:
iMarine Data Infrastructure Enabling Software contains the description of software and pointers to the documentation and artifacts of the related components that comprise the e-Infrastructure Management suite delivered from M11 to M24

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.	D8.3
Deliverable Title	iMarine Data Infrastructure Enabling Software
Contractual Delivery Date	31/10/2013
Actual Delivery Date	31/10/2013
Author(s)	Alex Antoniadis (NKUA), Roberto Cirillo (CNR), Ciro Formisano (E-IIS), Luigi Fortunati (CNR), John Gerbesiotis (NKUA), Lucio Lelii (CNR), Panagiotis Liakos (NKUA), Antonio Savini (E-IIS), Fabio Simeoni (FAO), Ermannno Travaglino (E-IIS).
Editor(s)	Luigi Fortunati (CNR)
Reviewer(s)	Andrea Manzi (CERN)
Contributor(s)	N/A
Work Package No.	WP 8
Work Package Title	iMarine Data e-Infrastructure Enabling-technology Development
Work Package Leader	Luigi Fortunati (CNR)
Work Package Participants	CNR, NKUA, FAO, E-IIS
Estimated Person Months	21,6
Distribution	Public
Nature	Other
Version / Revision	1.0
Draft / Final	Final
Total No. Pages (including cover)	12
Keywords	Enabling Software, Workflow, Process Execution Engine, Security, gCube, Maven

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
VO, Virtual Organization	A dynamic set of individuals or institutions defined around a set of resource-sharing rules and conditions. All these virtual organizations share some commonality among them, including common concerns and requirements, but may vary in size, scope, duration, sociology, and structure.
VRE, Virtual Research Environment	A *system* with the following distinguishing features: *(i)* it is a Web-based working environment; *(ii)* it is tailored to serve the needs of a Community of Practice; *(iii)* it is expected to provide a community of practice with the whole array of commodities needed to accomplish the community's goal(s); *(iv)* it is open and flexible with respect to the overall service offering and lifetime; and *(v)* it promotes fine-grained controlled sharing of both intermediate and final research results by guaranteeing ownership, provenance, and attribution.
Scope	Runtime information about the *Infrastructure*, the *Virtual Organization* (VO), or the *Virtual Research Environment* (VRE) in which resources of the infrastructure are operating in a specific moment of their lifetime.
XACML, *eXtensible Access Control Markup Language*	A policy-based language standardized by OASIS to make the use of security for software easy to configure, extend, and change with minimal impact.
gRS2, gCube Result Set 2	A framework, part of the gCube system, enabling point to point producer/consumer communication
Shibboleth	Shibboleth is a 'single-sign in', or logging-in system for computer networks and the Internet. It allows people to sign in, using just one 'identity', to various systems run by 'federations' of different organizations or institutions. The federations are often universities or public service organizations.
Identity Management System	Information system that can be used for enterprise or cross-network identity management.
SAML	SAML (Security Assertion Markup Language) is an Extensible Markup Language (XML) standard that allows a user to log on once for affiliated but separate Web sites. SAML is designed for business-to-business (B2B) and business-to-consumer (B2C) transactions.
GCube Hosting Node	The topological unit of a gCube infrastructure. An abstraction over a container running on a given port and hosting at least a minimal set of basic gCube services (the local services) dedicated to the host management.
Portal bundle	Archive that bundles an installable instance of the GCube infrastructure Gateway.

DELIVERABLE SUMMARY

1. INTRODUCTION

This deliverable describes the novelties and evolution of the iMarine Data Infrastructure Enabling Software from M11 (Sep '12) to M24 (Oct '13).

2. TARGET RELEASE(S)

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

- gCube 2.11 and gCube 2.11.1
- gCube 2.12
- gCube 2.13
- gCube 2.14
- gCube 2.15
- gCube 2.16 and gCube 2.16.1
- gCube 2.17

3. OBJECTIVES

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

a) T8.1 - Enabling-technology Development:
   a. Maven Support
   b. Featherweight Stack and SmartGears
b) T8.2 - Policy-oriented Security Facilities:
   a. Authentication and Identity Federation
   b. Accounting
c) T8.3 - Workflow Management Facilities:
   a. MapReduce compliance
   b. Workflow HiveQL Adaptor
   c. Workflow Orchestration Layer
   d. Workflow enhancements
   e. Resource Registry enhancements

1.3.1 T8.1 - ENABLING-TECHNOLOGY DEVELOPMENT

Maven support

As Maven started to be supported at the end of September 2012, WP8 focused on providing support to the other WPs by providing support tools helping developers on the development of gCube services and libraries capable of being integrated in the infrastructure.

Starting from September 2012 a component (maven-service-plugin) was released which allowed automatic generation of GAR (gCube Service Archive) artifacts for gCube Service maven projects. This allowed developers to be able to autonomously generate GARs using maven and deploy them manually in a gCube container sensibly simplifying testing procedure.

By supporting Maven on the build and integration infrastructure it was decided to develop a gCube Bill of Materials (BOM) component. This component can control the version of direct and indirect component dependencies and alert developers of dependency clashes with library versions included in the GHN and portal distribution bundle.

Featherweight stack and SmartGears

Starting with October 2012, interest began to grow in the area of Infrastructure exploitation by external clients. Effort has been put on providing a lighter version of GHN stack in order to deliver it to partners interested on leveraging gCube services. Before its distribution it was mandatory to bring into the client application the entire GHN server side stack in order for a client to interact with infrastructure services. A new mavenized version for the components that allows to invoke gCube services was released in October 2012, greatly reducing the size of the embedded stack on client side (ghn-client-runtime and ghn-client-notification). At the end of 2012, WP8 started the Featherweight Stack (FWS) initiative, a project with the aim to produce a set of tools that simplifies the connection to the infrastructure by external clients. The tools provided should be small and simple in terms of code, dependencies and overall usage requirements. A number of components of the FWS were released in December 2012 with the gCube 2.12 release, mainly addressing the area of Information System resources retrieval, along with a newer utility for efficient unit testing of gCube services (my-container). New components were later released in the first months of 2013 covering the area of resource publication and resource data encryption.

In order to implement those features, new technologies were used: Java JAX-WS allowed to build web services easily as proxies to gCube Services, JAXB annotated classes allowed to create a new Java Bean model mapping the existing Information System resource model.

The evolution path planned towards openness and simplification already started at the end of June 2013 when WP8 focused on working on the first implementation of a gCube integrated web service capable of being deployed on a container different than the GHN container.

Traditionally, enabling SaR (Software-as-a-Resource) and CaR (Container-as-a-Resource) has been the main value proposition of the gCube Core Framework (gCF). The way gCF delivers that value has limitations, however:

- "Closed world": gCF tightly ties developers to implement web services specifically for the gCube infrastructure. Services developed for the gCube system should be modified in order to be used in other containers different than the gHN.
- Technology constraints: The only container enabled is the Globus container and the only web service technology available is JAX-RPC

- Web Services are strongly coupled with gCF API, a change on the API can break compliance with implemented services, limiting the evolution of the system.
- Age: gCF, Globus, and their technological context are well dated by now
- Visibility: gCF sits right at the top of the software's stack and right in the middle of its design

In order to address these issues that can limit the evolution and the usability of the system it has been decided to develop a set of libraries under the name of *SmartGears* which can turn servlet based containers and web applications into gCube resources, transparently.

Since visibility was the key problem to address: if gCF wasn't visible to begin with, its age would be of little of no concern. With Smartgears we proposed the same net value as gCF, but deliver it in a completely different fashion: we moved away from frameworks and make Smartgears invisible to the software, not part of its stack at all. As a result, gCube is invisible too and any web application can run in the infrastructure: SaR becomes a nature that software acquires at runtime.

1.3.2   T8.2 - POLICY-ORIENTED SECURITY FACILITIES

Authentication and Identity Federation

The security Framework used by iMarine, SOA3, has been extended with two new functionalities: Distinguished Name based Authentication and SAML Federation. The Distinguished Name of the X509 Certificates used for HTTPS requests can be registered in the LDAP or in the gCube Information System. In the first case the DN is a specific user attribute and identifies a unique user, in the second case it is associated to a GHN instance and identifies the service instances running on that GHN. The module providing DN based Authentication is part of the SOA3 Connector and allows to query the Information System and the LDAP server for a profile associated to the DN: if a profile exists the user (or service) is authenticated, otherwise it is rejected [9].

The SAML Federation module uses Shibboleth Service Provider for authentication against an external Identity Management System: a successful authentication generates a SAML Assertion which, during its lifetime, is used to identify the caller inside gCube Infrastructure [10].

Lastly, a new filter has been added on the portal side in order to put the SAML Assertion ID in the User Session: this way the assertion can be used by service clients running inside the portal. The Integration with SAML federation has been completed by modifying the login page and by adding a *Discovery Service* allowing each user to select his/her home domain [11].

Authorization
The implementation of the Authorization Module has been completed and a Portlet for defining Authorization Policies has been developed [12]. The Portlet has been conceived to help the user to define Attribute based Policies to be used in iMarine context.

Attribute based Policies are composed of four attributes:

- *Subject*, defining who (or what) should be authorized or unauthorized

- *Action*, defining what the subject would like to do

- *Resource*, defining the target of the action

- *Environment*, defining under which conditions the action should be authorized or unauthorized

Accounting

In a production environment such as iMarine infrastructure it's important to keep track the consumption and allocation of the resources to users. For this reason Accounting and Billing are mandatory .

In order to provide those functionality in iMarine, the starting point was the Accounting & Billing system developed under the VENUS-C EU research project [13].

This system is built around the Usage Tracker service whose goal is to keep track of resource usage by receiving and archiving accounting records. It provides RESTful and Java APIs to perform operations on the accounting information.

For the Resource Accounting system [14] in iMarine the accounting record data-model [15] was defined to represent the information to collect and the Usage Tracker has been modified and enhanced to take into account a generic type of resource and to allow the infrastructure administrators to understand both the resources to account to a particular user and the usage trends.

To expose the accounting information to the users of the iMarine infrastructure, a Resource Accounting Portlet [17] has been designed and developed following the Visual Information-Seeking Mantra [18], showing summarized views at a first glance (accounting reports) and allowing the users to dive into more detailed data (accounting records).The accounting information are retrieved by the Portlet using the Usage Tracker API.

In order to allow the services to produce accounting information and to allow the Usage Tracker to consume those information the common-accounting library [15] has been designed and developed.

1.3.3 T8.3 - WORKFLOW MANAGEMENT FACILITIES

MapReduce Compliance

PE2ng has been refactored to comply with Map-Reduce / Master-Worker model abstraction. A library of operators which can be used in order to define Map-Reduce jobs has been developed. Every Map/Reduce job will be materialized as an ordinary execution plan of PE2ng. Those plans will be consisted of PE2ng Operators, which should be considered as the building blocks of every Map/Reduce job workflow.

Workflow HiveQL Adaptor

A new Workflow Engine adaptor was implemented, capable of parsing HiveQL queries and exploiting the recently introduced MapReduce related operators. It allows the submission of jobs that comply with Map-Reduce / Master-Worker parallel execution model abstractions, while supporting various sources and output stores. This way, no staging of data is required, as it is not necessary to provide the input neither retrieval of the execution output, because this is done automatically.

Workflow Orchestration Layer

The Workflow Orchestration Layer provides a high level interface which enables the description and handling of workflows executions over heterogeneous processing infrastructures (PE2NG, Grid, Hadoop, Condor). PE2NG was extended in order to process web service calls by means of an extension of JDL,

namely gJDL, which allows the description of additional job types. Through this layer, an abstract plan can be submitted in order to be executed in the available underlying infrastructures.

Workflow Enhancements

Various enhancements were introduced to the Workflow layer:

- Grid Adaptor was able to handle execution of JDL jobs consisted of a single task, but now can carry out jobs with multiple sub-tasks.
- Node selection was improved. Now, timestamps have better precision. Also, during a tie break procedure all available nodes are taken into account, even those not selected in the past.
- JDL adaptor performance got improved, as it now organizes file transfers to happen concurrently. When many file transfers are needed, they get organized in small groups.

Resource Registry Enhancements

Resource Registry has been extended to support reset operation in order to recover from internal failovers or network failures that might leave its internal database in an inconsistent state and avoid container restart. Moreover, another extension has been developed to support a per Virtual Organization (VO) configuration. Moreover, Host Nodes filtering based on some user-specific criteria functionality has been added in order to provide the ability to limit the available hosting nodes, depending on their installed libraries, in a plan creation scenario.

4. COMPONENTS

Maven support:

- org.gcube.build-utils.maven-parent-1.0.0
- org.gcube.distribution.maven-bom.1-3-0
- org.gcube.distribution.maven-portal-bom.1-3-0

Featherweight Stack:

- common-gcore-stubs 1.1.0
- common-gcore-resources 1.2.0
- discovery-client 1.0.1
- ic-client-1.0.1
- registry-publisher 1.2.1
- common-scope 1.2.0
- common-scope-maps 1.0.1
- common-configuration-scanner 1.0.0

Smartgears:

- common-events-1.0.0
- common-smartgears-1.0.0
- common-validator-1.0.0

- common-smartgears-app-1.0.0
- maven-smartgears-bom 1.0.0
- smartgears-distribution 1.0.0

Federated Authentication:

- org.gcube.vo-management.idfederation.samlconsumer.0-1-0
- org.gcube.vo-management.soa3.authentication.rest.0-6-0

Authorization:

- org.gcube.vo-management.soa3.authorization.policymanagementAPIimplementation.0-1-0
- org.gcube.vo-management.soa3.authorization.policymanagementAPIinterfaces.0-1-0
- org.gcube.vo-management.soa3.authorization.policymanagementinterface.0-5-0
- org.gcube.vo-management.soa3.authorization.policymanagementservice.0-1-0
- org.gcube.portlets-admin.policy-definition-portlet.2-0-0

Distinguished Name Authentication:

- org.gcube.vo-management.soa3.connector-message-security-manager.0-5-0
- org.gcube.vo-management.soa3.connector.0-5-0
- org.gcube.vo-management.connector.gss.integration.0-5-0
- org.gcube.vo-management.soa3.connector.common-security.0-5-0
- org.gcube.vo-management.soa3.connector.integrationlib.0-5-0
- org.gcube.vo-management.soa3.connector.service.0-5-0
- org.gcube.vo-management.soa3.connector.utils.0-5-0
- org.gcube.vo-management.soa3.usermanagement.0-5-0
- org.gcube.vo-management.soa3.usermanagement.rest.0-5-0

Resource Accounting:

- org.gcube.accounting.common 2.0.0
- org.gcube.accounting.usagetracker 2.0.0
- org.gcube.portlets.admin.accounting-portlet 1.0.0
- org.gcube.accounting.common-accounting-lib 1.0.0
- org.gcube.accounting.common-accounting-model 1.0.0
- org.gcube.vo-management.soa3.authentication.rest.0-6-0

Map reduce compliance:

- org.gcube.search.operatorlibrary.1-3-0data-handlers 1.0.0

Workflow enhancements:

- org.gcube.execution.WorkflowEngine 1.3.1-1.3.2
- org.gcube.execution.MadgikCommons 1.4.0

Workflow HiveQL Adaptor:

- org.gcube.execution.WorkflowHiveQLAdaptor 1.0.0
- org.gcube.search.data-handlers-library 1.0.0
- org.gcube.search.hive-parser 1-0-0

Workflow Orchestration Layer:

- org.gcube.execution.WorkflowEngine 1.3.3
- org.gcube.execution.WorkflowEngineClientLibrary 1.0.0
- org.gcube.execution.WorkflowOrchestrationLayerService 1.0.0
- org.gcube.execution.WorkflowOrchestrationLayer-Client 1.0.0

Resource registry enhancements:

- org.gcube.execution.RRGCubeBridge.1-4-0
- org.gcube.execution.ResourceRegistry_NODEPS.1-4-0
- org.gcube.execution.ResourceRegistry.1-4-0
- org.gcube.execution.RRGCubeBrigde.1-5-0

5. DOCUMENTATION

A comprehensive overview of the subsystems the described components belong to is available at [6] (e-infrastructure Management Facilities), [7] (Policy-oriented Security Facilities) and [8] (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]

For development purpose, Javadoc documentation for each component, along with a direct link to the associated section in Developer's Guide, is available at [1].

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 34: Data e-Infrastructure Policy-oriented Security Facilities:
       https://gcube.wiki.gcube-system.org/gcube/index.php/Data_e-Infrastructure_Policy-oriented_Security_Facilities

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

[9]   DN Based Authentication:
       https://gcube.wiki.gcube-system.org/gcube/index.php/DN_Based_Authentication

[10]   SOA3 Authentication Service:
       https://gcube.wiki.gcube-system.org/gcube/index.php/SOA3_Authentication_Service

[11]   Shibboleth and GCube:
       https://gcube.wiki.gcube-system.org/gcube/index.php/Shibboleth_and_gCube

[12]   Policy Management Portlet:
       https://gcube.wiki.gcube-system.org/gcube/index.php/Policy_Management_Portlet

[13]   Venus-C Project home page:
       http://www.venus-c.eu/Pages/Home.aspx

[14]   Resource Accounting:
       https://gcube.wiki.gcube-system.org/gcube/index.php/Resource_Accounting

[15]   Accounting model:
       https://gcube.wiki.gcube-system.org/gcube/index.php/Common-accounting-model

[16]   Usage Tracker:
       https://gcube.wiki.gcube-system.org/gcube/index.php/Usage_Tracker

[17]   Accounting Portlet:
       https://gcube.wiki.gcube-system.org/gcube/index.php/Accounting_Portlet

[18]   B. Shneiderman: "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations". IEEE Symposium on Visual Languages, IEEE (1996)