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 e-Infrastructure Operation Report*
Deliverable No.	**D5.3**
Delivery Date	***June 2013***
Author	***Andrea Manzi (CERN), Leonardo Candela (CNR)***

Abstract: *The iMarine Data e-Infrastructure is a living system whose development is mainly driven by requirements and feedback produced by the iMarine CoP. This deliverable is the second of the series of reports on the activities performed while operating such a system and describes the status of the Data e-Infrastructure in terms of nodes available, software deployed, quality of the service, and usage.*

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.	D5.3
Deliverable Title	iMarine Data e-Infrastructure Operation Report
Contractual Delivery Date	June 2013
Actual Delivery Date	June 2013
Author(s)	Andrea Manzi, CERN, Leonardo Candela, CNR
Editor(s)	Andrea Manzi, CERN
Reviewer(s)	George Kakaletris, NKUA
Contributor(s)	Kostantinos Kakaletris (NKUA), Anton Ellenbroek (FAO), Tommaso Piccioli (CNR)
Work Package No.	WP 5
Work Package Title	iMarine Data E-Infrastructure Deployment and Operation
Work Package Leader	Andre Manzi (CERN)
Work Package Participants	CERN, CNR, NKUA, FAO, FIN
Estimated Person Months	11,845
Distribution	Public
Nature	Report
Version / Revision	1.0
Draft / Final	Final
Total No. Pages (including cover)	41
Keywords	Infrastructure Operation; Federated Resources; Virtual Organisation;

CHANGE LOG

Reason for Change	Issue	Version	Date	Partner / Responsible
First TOC		0.1	15/05/2013	CERN
Reviewed TOC		0.2	21/05/2013	NKUA/CERN
First version ready for internal review (Section 2 to be updated by CNR)		0.3	05/06/2013	CERN
First version of Section 2. Minor changes.		0.4	06/07/2013	CNR
Applied some changes after a first internal review		0.5	07/06/2013	CERN
Version with Reviewer comments		0.5.1	27/06/2013	NKUA
Version addressing Reviewer comments		0.6	28/06/2013	CERN
Final version		1.0	30/06/2013	CERN/NKUA

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
CoP (Community of Practice)	A term coined to capture an "activity system" that includes individuals who are united in action and in the meaning that "action" has for them and for the larger collective. The communities of practice are "virtual", i.e., they are not formal structures, such as departments or project teams. Instead, these communities exist in the minds of their members, are glued together by the connections they have with each other, as well as by their specific shared problems or areas of interest. The generation of knowledge in communities of practice occurs when people participate in problem solving and share the knowledge necessary to solve the problems.
DPM	Disk Pool Manager
EAF	Ecosystem Approach to Fisheries
e-Infrastructure	An operational combination of digital technologies (hardware and software), resources (data and services), communications (protocols, access rights and networks), and the people and organizational structures needed to support research efforts and collaboration in the large.
GHN	gCube hosting Node
iMarine	Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources.
LFC	LGC File Catalog
LME	Large Marine Ecosystem
NGI	National Grid Initiative
QA	Quality Assurance
sBDII	Site BDII
Virtual Research Environment (VRE)	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.
VOMS	Virtual Organization Membership service
Virtual Organization (VO)	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.
WMS	Workload Management System

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The objective of the "**iMarine Data e-Infrastructure Deployment and Operation**" work package is to effectively deploy and maintain the computing resources shared in the iMarine Data e-Infrastructure and make available the software and tools (infrastructure administration portal) for WP6 to operate the Virtual Organisations (VOs) and Virtual Research Environments (VREs) that serve the user communities' needs. The Data e-Infrastructure deployed during the iMarine project, contributes and takes part of the D4Science Infrastructure, a data infrastructures maintained by the D4Science initiative [1]. Such an ecosystem is a result from the joint effort and resources of a number of projects including iMarine [2], EUBrazilOpenBio [3], and ENVRI [4]. In particular, the iMarine project coordinates and operates the Virtual Organisations, the Virtual Research Environments and the resources needed to serve the needs of the iMarine CoP [5].

This report documents the work done to operate the iMarine resources part of the D4Science infrastructure during the period from August 2012 to May 2013, which corresponds to months M10 to M19 of the iMarine project. That is, it describes the VOs and VREs deployed to satisfy the iMarine CoP needs.

From the perspective of the overall service offered to the end user, different environments have been maintained, deployed and made available, satisfying the requirements expressed by the different project scientific clusters. In total 2 VOs and 14 VREs were provided. The management of the infrastructure was facilitated by the implementation of straightforward procedures for monitoring and accounting that have been defined during the first period of the project. A number of monitoring tools was put in place and subsequently was enhanced to allow infrastructure managers to visualize the status of the resources and to be actively notified when problems occurred. An accounting tool provides relevant statistics about the users' exploitation of the infrastructure and the infrastructure load. A straightforward support procedure continued to be implemented during the course of this project months. A total number 149 tickets were submitted, 56.3% being of high priority. All tickets were properly closed and documented.

This report, after introducing its the goal and concepts, (a) enumerates the set of Virtual Organisations and Virtual Research Environments operated, (b) presents the activities related to the deployment and operation of the infrastructure (c) discusses the facilities for monitoring and accounting and presents derived indicators and finally presents the activity performed by the production support team to deal with issues that rose during the operation of the infrastructure.

1 INTRODUCTION

This deliverable describes (a) the iMarine resources that compose the D4Science Infrastructure, and (b) the activities carried out during the second period of the iMarine project to operate such resources. Before describing the resources which compose the infrastructure, it's important to highlight the main aspect that characterizes the infrastructure. The D4science Infrastructure is composed by resources (computational, storage, data), which are collectively labelled as iMarine Resources. Those resources can be shared among the users of a Virtual Organization (VO). Finally, a Virtual Research Environment (VRE) aggregates and deploys on-demand content resources and application services by exploiting computational and storage resources of grid and cloud infrastructures belonging to a VO.

The D4Science Infrastructure is organized in three resource types, hosted by the project partners:

- **gCube resources** are the physical or virtualized resources running gCube software (the gCube Hosting Node) that makes them capable running gCube services or libraries providing the functionality to create, manage, and exploit VREs;

- **UMD resources** are computing and storage nodes running the UMD software[8]. UMD is the result of the integration and certification process (performed by the EGI-INSPIRE project) of the EMI middleware provided by the EMI project[9]. By running UMD, these nodes provide core grid functionalities such as file-based storage, distributed computation of applications, etc. EMI nodes are exploited by gCube services which then provide higher level functionality through the iMarine VREs.

- **Runtime resources** are third party services/software exploited at runtime by the iMarine VREs. The Runtime Resources are modelled as a particular type of resources on the Information System giving the possibility to the gCube services to dynamically discover their location and their encrypted access information.

In the previous deliverable of the series D5.2 [18], is describing the aforementioned concepts in detail.

In addition to the D4Science infrastructure resources which are managed by the project partners, users can have access to resources accessible following a federated approach. Therefore this report covers both hosted and federated types of resources.

During the second period of the project the focus was on consolidating and enhancing the resources needed by iMarine VOs and VREs. One VO (Ecosystem) which was part of the infrastructure during the first period, has been dismissed in order to reassign the related resources to QA processes (see section 3.3) and enlarge the 2 production VOs which are accessible to iMarine users (gCubeApps and FARM VOs).

The rest of this report is organised as follows.

Section 1 introduces the document and its goal.

Section 2 describes the set of Virtual Organisations and Virtual Research Environments operated to serve the needs of the iMarine CoP.

Section 3 presents the activities related to the deployment and operation of the infrastructure in terms of its resources including hosted and federated resources.

Section 4 discusses the facilities for monitoring and accounting as well as presents some indicators resulting from their exploitation.

Section 5 presents the activity performed by the production support team to deal with issues and malfunctions raised by CoP members while exploiting the Infrastructure and its services.

2 VIRTUAL ORGANIZATION AND VIRTUAL RESEARCH ENVIRONMENTS

Virtual Organisations (VOs) and Virtual Research Environments (VREs) are, from the infrastructure operation point of view, sets of resources and users grouped together by sharing policies with the goal to serve the needs of a certain case.

The set of iMarine VOs and VREs the D4science infrastructure operates, as well as their evolution in terms of resources involved and services offered, is mainly a consequence of the requirements captured in the context of the 3 iMarine scientific business cases the project is focused on:

1. Support to EU Common Fisheries Policies;
2. Support to FAO's deep seas fisheries programme: balancing use of marine resources and protection of vulnerable marine ecosystems in the high seas;
3. Support to regional (Africa) LME pelagic EAF community.

These requirements are documented in "Ecosystem Approach Community of Practice: Requirements" [6]. Specified requirements are carefully analysed and transformed into a development plan as documented in "Virtual Research Environment Deployment and Operation: Plan" [7]. This development plan is then implemented by a dedicated team and a Wiki page is updated to document the activities performed. In addition to the 3 business cases introduced above, other VOs and VREs have been maintained from the previous projects Communities[1] to serve the needs of "external" communities having expressed their interest in these services.

This section provides a brief report on the VOs and VREs deployed and maintained by the iMarine project up to M19, i.e. May 2013.

2.1 VIRTUAL ORGANISATIONS (VOS)

During this reporting period 2 iMarine Virtual Organisations have been maintained and operated in the context of the D4Science infrastructure:

- The **FARM** Virtual Organisation, created for the Fisheries and Aquaculture Resources Management communities. This VO supports a large number of application scenarios from these communities such as the production of Fisheries and Aquaculture Country Profiles, the management of catch statistics including the harmonisation across data-sources, the dynamic generation of biodiversity maps and species probability maps, the analysis of vessel trajectories. It is VO candidate to host the applications and services realising scenarios identified by the iMarine CoP;
- The **gCubeApps** Virtual Organisation, created to host a number of VREs focusing on specific applications ranging from ecological niche modelling to time series management and vessel trajectories analysis.

The **Ecosystem** VO was part of the production environment up to April 2013. Since (a) this VO was created to host Virtual Research Environments that are not in the scope of iMarine and (b) the activities the Ecosystem's VREs were conceived for are either no longer active or have been assigned to other VREs, it was decided to use this VO as a pre-production environment. This makes it possible for the WP5 team to perform tasks towards enhancing the quality of the services offered via the other VOs.

[1] The D4science and D4science II projects User Communities

VO	VREs	Users
FARM	7	76
gCubeApps	7	76

Table 1 - iMarine VOs detailed information

2.2 VIRTUAL RESEARCH ENVIRONMENT (VRES)

Fourteen (14) Virtual Research Environments have been implemented, each in the context of a specific Virtual Organisation.

2.2.1 THE FARM VRES

The FARM Virtual Organisation hosts and operates the following seven (7) Virtual Research Environments:

- The **AquaMaps** Virtual Research Environment is for providing fisheries and aquaculture scientists with facilities for producing and accessing species predictive distribution maps showing the likelihood that a certain species or a combination of species will live in specific regions or areas;
- The **Fisheries Country Profiles Production System (FCPPS)** Virtual Research Environment is for fisheries and aquaculture authors, managers and researchers who produce reports containing country-level data. It provides seamless access to multiple data sources, including their annotation and versioning and permits production of structured text, tables, charts and graphs from these sources to be easily inserted into custom reporting templates that can support multiple output formats;
- The **FishFinder (FishFinderVRE)** Virtual Research Environment is established to elaborate Species Fact Sheets, fill / view their metadata, and select data for download and-or display in a Stand-alone version of the VRE. The explicit purpose is to enable some 50 authors to prepare hundreds of species fact sheets;
- The **Integrated Capture Information System (ICIS)** Virtual Research Environment offers fisheries statisticians a set of tools to manage their data. Statisticians produce statistics from often very different data sources, and need a controlled process for the ingestion, validation, transformation, comparison and exploitation of statistical data for the fisheries captures domain;
- The **iMarine Board (iMarineBoard)** Virtual Research Environment is designed to provide the members of the iMarine Board with collaboration tools and a demonstration of infrastructure facilities. This VRE therefore includes those services that put into effect iMarine governance models and policies, such as (i) the collaboration suite including a shared workspace and messaging system, (ii) services for accessing biodiversity data from several major databases, and (iii) services for managing tabular data (e.g. catch statistics) and code lists;
- The **Vessel Transmitted Information (VTI)** Virtual Research Environment is for marine biologists willing to analyse vessel activities over space and time by taking into account environmental data;
- The **Vulnerable Marine Ecosystem Database (VME-DB)** Virtual Research Environment is for fisheries and aquaculture authors willing to collaboratively produce Fact Sheets on Vulnerable Marine Ecosystems (VME).

2.2.2 THE GCUBEAPPS VRES

The gCubeApps Virtual Organisation hosts and operates the following seven (7) Virtual Research Environments:

- The **BiodiversityResearchEnvironent** Virtual Research Environment is conceived to provide biodiversity scientist with facilities for seamless access to a rich array of biodiversity data including occurrence points and taxa records from established providers including GBIF, Catalogue of Life, and OBIS.
- The **DocumentsWorkflow** Virtual Research Environment is conceived to provide its users with a working environment focused on the gCube facilities for managing Document life-cycles. It exploit the facilities offered by the gCube Business Documents Workflow Management Suite enabling the production of reports that require a collaborative activity of several actors;
- The **EcologicalModeling** Virtual Research Environment is conceived to provide its users with a working environment focused on the gCube facilities for producing species distribution maps resulting from the processing of data on species characteristics and environmental observations. The resulting maps are actually rich information objects containing PNG images, GIS layers as well as metadata;
- The **Scalable Data Mining** Virtual Research Environment is conceived to cater for Data Mining techniques to be applied to biological data. The algorithms are executed in a distributed fashion on the e-Infrastructure nodes or on local multi-core machines. Statistical data processing can be applied to perform Niche Modelling or Ecological Modelling experiments. Other applications can use general purpose techniques like Bayesian models. Time series of observations can be managed as well, in order to classify trends, catch anomaly patterns and perform simulations.
- The **TCom** Virtual Research Environment is for the members of the iMarine Technical Committee. In essence it provides the members of this committee with a working environment based on gCube services.
- The **TimeSeries** Virtual Research Environment is conceived to provide its users with a working environment focused on gCube facilities for managing time series. This environment supports the load of time series objects, the curation and validation by relying on authoritative code lists, the sharing of such objects with co-workers, the production of graphs, the visualization through a GIS service;
- The **VesselActivitiesAnalyzer** Virtual Research Environment is conceived to provide its users with a working environment focused on gCube facilities for managing vessel trajectories. This environment support users in loading and curating their own vessel trajectories, enriching such data with bathymetry and FAO Area, sharing with co-workers, analysing such objects by producing maps on vessel activities and fishing monthly effort;

2.2.3 INDICATORS ON CURRENT VRES AND THEIR EVOLUTION

VRE	VO	Creation date[2]	Last update	Users[3]
AquaMaps	FARM	M1, Nov 2011	M19, Jun 2013	54
FCPPS	FARM	M1, Nov 2011	M19, Jun 2013	38
FishFinderVRE	FARM	M16, Mar 2013	M19, Jun 2013	9
ICIS	FARM	M1, Nov 2011	M19, Jun 2013	47

[2] With M1 we highlight the fact that this VRE was created in the context of previous projects and maintained by iMarine.

[3] At the date of writing this report, i.e. June 2013.

iMarineBoardVRE	FARM	M15, Feb 2013	M19, Jun 2013	13
VME-DB	FARM	M9, Jul 2012	M19, Jun 2013	11
VTI	FARM	M1, Nov 2011	M19, Jun 2013	25
BiodiversityResearchEnvironment	gCubeApps	M9, July 2012	M19, Jun 2013	41
DocumentsWorkflow	gCubeApps	M1, Nov 2011	M19, Jun 2013	31
EcologicalModeling	gCubeApps	M1, Nov 2011	M19, Jun 2013	45
ScalableDataMining	gCubeApps	M17, Oct 2012	M19, Jun 2013	26
TCom	gCubeApps	M17, Apr 2013	M19, Jun 2013	28
TimeSeries	gCubeApps	M1, Nov 2011	M19, Jun 2013	36
VesselActivitiesAnalyzer	gCubeApps	M1, Nov 2011	M19, Jun 2013	33

Table 2 - iMarine VREs detailed information

The following three figures describe the evolution in term of users served by a given VRE with respect to the previous reporting period [18]. In particular, Figure 1 describes the evolution of the VREs operated by the FARM VO while Figure 2 describes the evolution in the case of the gCubeApps VO.

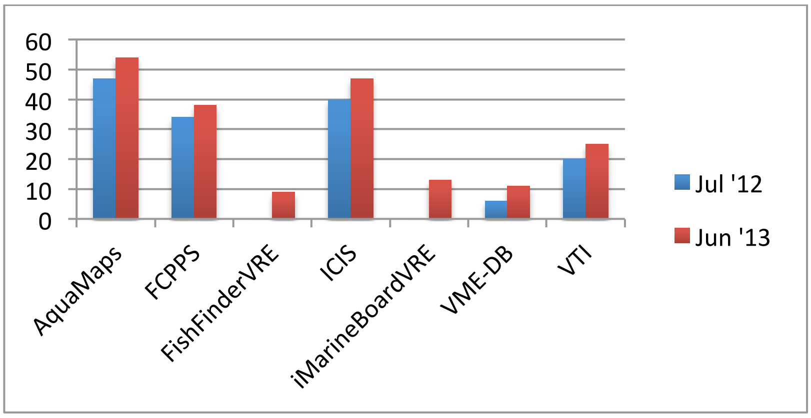

Figure 1. FARM VO: Evolution of number of users per VRE

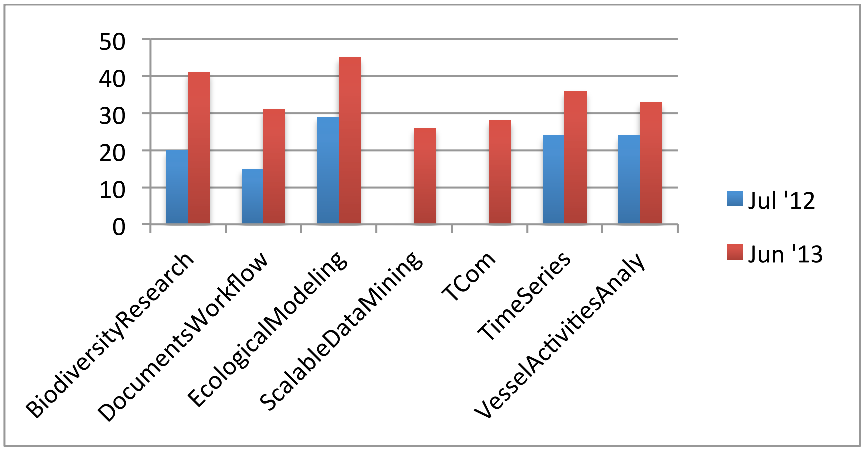

Figure 2. gCubeApps VO: Evolution of number of users per VRE

Figure 3 describes the evolution by using the total number of users served by the VREs of the two VOs. In average, it has been observed an increase of +34% (147 at Jul ’12 vs 197 at Jun ‘13) in the number of users served by FARM VREs and of +114% (112 at Jul ’12 vs 240 at Jun ‘13) in the number of users served by gCubeApps VREs.

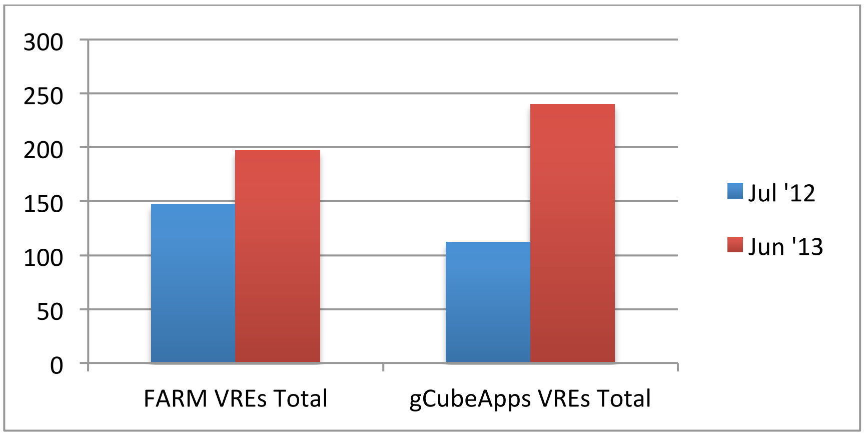

Figure 3. Evolution of total number of users served by VREs in a given VO

2.3 GATEWAYS

From the access point of view, the two gateways previously deployed have been maintained:

- **iMarine gateway** [https://portal.i-marine.d4science.org/]: it provides access to the FARM VO and its VREs and the gCubeApps VO and its VREs;
- **D4science.org gateway** [https://portal.d4science.org/]: it provides access to the Ecosystem VO and its VREs and to the gCubeApps VO as well.

The two gateways are the entry points to the infrastructure, offering not only a web friendly interface over the infrastructure service, but real complex environments accessible by the iMarine users through a thin layer (i.e. the browser).

3 RESOURCES OPERATION

As mentioned in the previous section, one of the goals of this deliverable is to describe the activities carried out during the second period of the project lifetime to operate the iMarine resources of the D4Science infrastructure. The following sections are meant to detail two groups of resources:

- Hosted Resources (cf. Section 3.1), i.e. resources deployed by the iMarine project partners and thus owned by the D4Science infrastructure;
- Federated Resources (cf. Section 3.2), i.e. resources deployed to serve the needs of existing projects and initiatives and federated by the D4Science infrastructure.

In addition we detail the activities carried out to perform gCube resources deployment and upgrade (cf. Section 3.1.2.1) in order to maintain the infrastructure operational and the activities and resources dedicated to Quality & Assurance tasks (cf. Section 3.3).

3.1 HOSTED RESOURCES

As originally defined in deliverable D5.1 iMarine Data e-Infrastructure Plan [11], the D4Science infrastructure hosting resources dedicated to iMarine are provided by T5.2 tasks members plus an external partner (ASGC Taiwan). During the period of the project another partner has started contributing to the D4Science Infrastructure, VLIZ which is under the umbrella of UNESCO partner. In addition, the FIN node hosting one of the infrastructure nodes has left the production infrastructure. Therefore in total 4 project member sites contribute to the infrastructure:

- **CNR** – Pisa, Italy
- **FAO** – Rome, Italy
- **NKUA** – Athens, Greece
- **VLIZ** – Ostende,Belgium

In addition an external project partner is also providing resources:

- **ASGC** – Taiwan

As already mentioned, the FIN site that was part of the production infrastructure during the first period has been dismissed.

3.1.1 HOSTING RESOURCES

The following table provides detailed information about the contribution from each site ordered by CPUs. The table reports information about hosting hardware and virtualized resources. For this reason the column type either reports **Hardware** together with type of CPU or **VM** and the type of virtualization system.

Site	Type	RAM (GB)	Disk (TB)	CPUs
CNR	VM / Xen hypervisor	844	35.4	754
	Subtotal	844	35.4	754
NKUA	VM / Xen hypervisor	106.25	1.8	56
	Subtotal	106.25	1.8	56
ASGC	Hardware/ Two Quad-Core Intel(R) Xeon(R) CPU 5130 @ 2.00GHz	8	0.2	8
	Hardware/ Quad-Core Intel(R) Xeon(R) CPU 5150 @ 2.66GHz	4	0.1	4
	Subtotal	12	0.3	12
FAO	Hardware/Two Quad-Core Intel Xeon X5450 @ 3.0 GHz	8	0.3	8
	Subtotal	8	0.3	8
VLIZ	Hardware/Intel(R) Xeon(R) CPU E5649 @ 2.53GHz	3	0.4	4
	Subtotal	3	0.4	4
Total		**971.25**	**38.2**	**834**

Table 3 - Hosting resources by partner

The hosting resources deployed so far are compliant with the plans defined at the beginning of the project and allowed the deployment and availability of all VOs and VREs requested so far by the project.

3.1.2 GCUBE RESOURCES

The iMarine gCube resources hosted in the D4Science infrastructure are composed by multiple components that run on a special web service container, called gCube Hosting Node (gHN)[4]. Each gHN is configured to support one or more VO and can host one or more gCube services. These services can provide VO-level or VRE-level functionality.

The following table presents the number of gCube hosting nodes deployed per site :

Partner	gCube Hosting Nodes
CNR	57
NKUA	20
ASGC	3
FAO	2
VLIZ	1
Total	83

Table 4 - gCube Hosting Node per site

As said each hosting node can be assigned to one or more VOs. The following graph summarizes the gCube hosting nodes available for each VO. It includes also hosting nodes running gCube services in the "root" VO and hosting nodes not used at the moment ("spare"). The "root" VO hosts services which are at the base of the infrastructure functionality: the root instance of the Information System (where all GHNs are registered), the Software Gateway, the Messaging components etc.). Therefore it's accessible only by the Infrastructure Administrators and it does not group end-users service/functionalities. As shown in the graph a large part of the infrastructure nodes has been assigned so far to the gCubeApps VO, this is due to the computational requirement of the users of this VO where CPU intensive Data Mining computations are distributed among the VO resources.

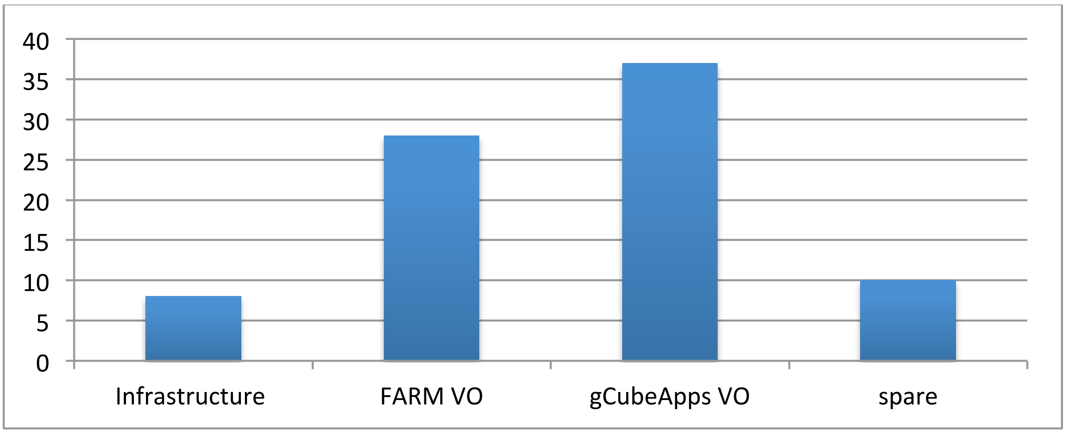

Figure 4 - gCube Hosting Node per VO

3.1.2.1 GCUBE DEPLOYMENT

[4] https://gcore.wiki.gcube-system.org/gCube/index.php/Main_Page

Currently, the latest gCube release placed in production is the release v2.14.0, deployed at the end of project's M19. Previously during the second project period, the infrastructure operation was based on 3 minor gCube releases (v2.11.0, v2.12.0 and v2.13.0). Due to the need to quickly fix critical defects affecting the production environment 2 gCube maintenance releases were also deployed (v2.9.1 and v2.11.1). provides more information about the gCube releases size.

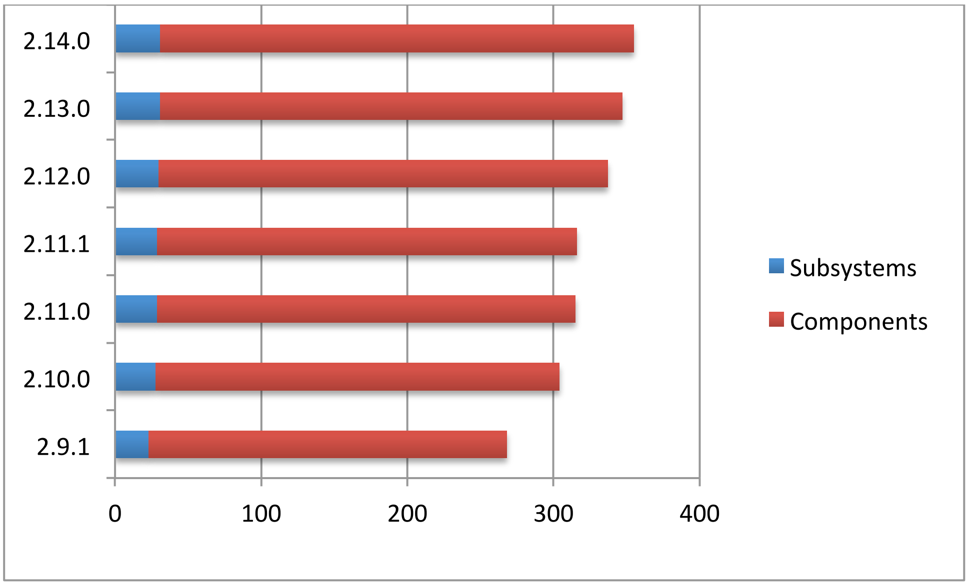

Figure 5 - gCube Releases Size

Starting from release v2.10.0 an increasingly new number of components have been integrated in the release cycle and therefore deployed in production. As reported the v2.10.0 release was not deployed in production immediately but the components integrated during that release cycle went to production during the upgrade to release v2.11.0. In addition the components released in v2.14.0 have been partially deployed in production at the end of M19. The rest of the upgrades plan to enter the production with the upgrades coming from release v2.15.0 (scheduled during M20).

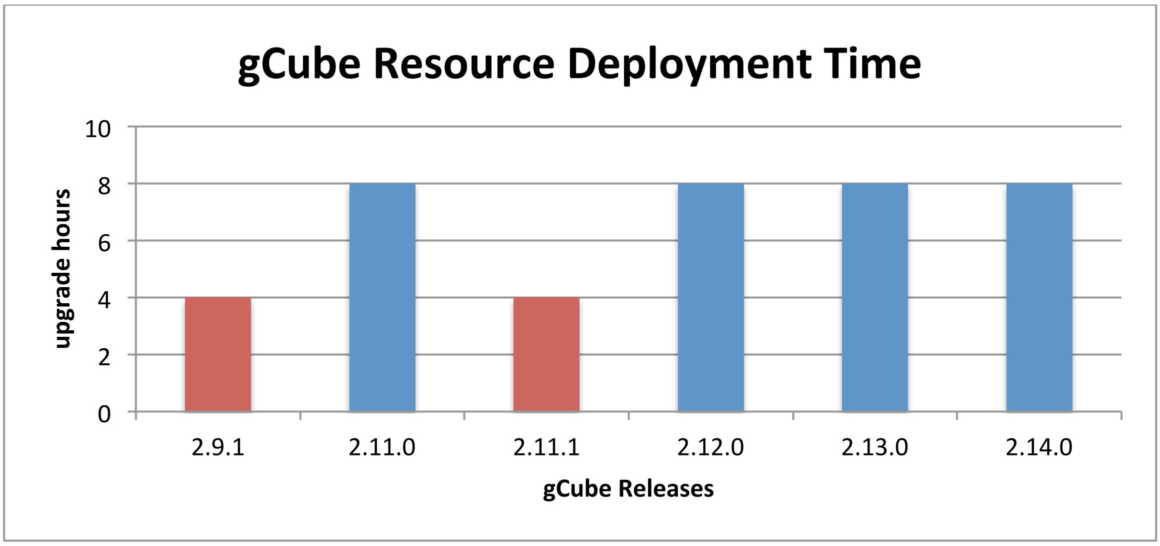

Figure 6 - gCube Releases deployment time

Figure 6 shows the number of hours needed to deploy any new gCube release in the infrastructure. The deployment of every minor release was scheduled and effectively required 8 hours while the deployment of a maintenance release required 4 hours. It should be noted that usually the upgrade to a new minor release requires modifications to all infrastructure nodes. Respect to the average upgrade time for a minor release during the first period of the project (20 hours) the parallelization of some upgrade tasks (infrastructure and gateways) has for sure enhanced the user experience.

3.1.3 UMD RESOURCES

Two sites (out of the four project sites) offer UMD resources to the infrastructure. UMD resources are computing and storage services running UMD middleware [8], which is a distribution for EGI of the software developed by the EMI project [9]. The UMD services are exploited by gCube services, which then provide higher-level functionality through the iMarine VREs. During the second period of the project, our sites running UMD version 1 have been upgraded to run UMD version 2. The upgrade was needed cause the UMD 1 services finished their support time at the end of February 2013.

The following table presents how the different UMD services have been distributed between the sites.

	CREAM	WN	sBDII	DPM	WMS	LFC	VOMS	LB	APEL
CNR	☑	☑	☑	☑	☑	☑	☑	☑	-
NKUA	☑	☑	☑	☑	☑	-	-	☑	☑

Table 5 - UMD services per Site

All UMD services provided by iMarine partners are also registered to join the EGI production infrastructure (cf. Section 3.2.1).

3.1.4 RUNTIME RESOURCES

This section describes the Runtime Resources which are deployed on the D4Science Infrastructure and exploited by the iMarine VREs components. The resources are grouped by their functional category (Third Party Services, Database, etc.).

THIRD PARTY SERVICES

The infrastructure hosts Services of various natures remotely accessible by the deployed gCube components.

The following table summarizes the Services that are deployed in the infrastructure.

Site	Service	Description	N of Instances
CNR	JackRabbit Repository	A JackRabbit repository is used as a backend for User workspace data.	1
CNR	ActiveMQ Message Broker	The ActiveMQ broker is the key point of both Monitoring and Accounting functionalities in iMarine, because it routes messages coming from different probes to consumers which then make available information about gHNs status, infrastructure usage and user exploitation of the iMarine portals.	1
CNR	Geoserver	The GeoServer service is deployed in the D4science infrastructure to enable GIS visualization of Species distribution and biodiversity maps. The Service is deployed together with an instance of PostgreSQL DB, in order to store GIS data.	4
CNR	RStudio	The R-Studio service is integrated within the gCube TimeSeries Environment to enable statistical elaboration of TimeSeries using the R language	1
CNR	SDMX Registry	The SDMX Registry allows the publication and registration of SDMX Data Sources. It's exploited by the TimeSeries Environment.	1

CNR	Thredds	The Thredds Service is responsible for the publication and discovery of geospatial data by means of OGC standard protocols	1
Terradue[5]	North52 WPS	The North52 WPS service implements standard geo-spatial processing interface. It has been integrated in iMarine with the Hadoop backend in order to increase the service scalability	1
CNR	Nexus Server	Nexus hosts the Maven Repositories containing gCube artefacts	1

Table 6 - Third Party Services deployed on the infrastructure

During the reported period, new developments required the deployment in production of the SDMX Registry ,Thredds and North52 WPS services.

DATABASES

External Databases are exploited by many gCube Services to persist various type of information ranging from TimeSeries, Codelists, Accounting & Monitoring data and GIS Layers. The following table summarizes details about the databases hosted in the infrastructure:

Site	Database Type	Description	N of Instances
CNR	PostgreSQL	The PostgreSQL DB backend for the gCube TimeSeries and Aquamaps Services	4
CNR	PostgreSQL	The PostgreSQL DB containing conversion tables between Fish Codes	1
CNR	PostgresSQL	The PostgreSQL DB backend hosting a mirror of the FishBase and SeaLifeBase DBs	1
NKUA	PostgresSQL	The PostgresSQL DB hosting the FLORA biodiversity repository	1

[5] The current production instance of WPS is hosted temporarily at Terradue premises. It will be moved during the last project period to one of the other WP5 project partners.

CNR	MySQL	The MySQL DB hosting the ICTV biodiversity repository	1
VLIZ	PostgreSQL	The PostgreSQL DB hosting the OBIS biodiversity Database	1
CNR	MySQL	The MySQL DB storing the D4Science Infrastructure Accounting and Monitoring data.	1

Table 7 - Databases deployed on the infra

It's important to highlight that during the second period of the project the infrastructure started to host one of the mirror of 2 important DBs like SeaLifeBase[6] and FishBase[7]. The info coming from those DBs are also exploited by the iMarine mobile app AppliFish[8].

CLUSTERS

This category of Runtime Resources groups together clusters of resources exploited either for computation or storage purposes by gCube components. This type of resources has the characteristic to be elastic and transparent extensible in terms of number of nodes. The following table give details on the cluster deployed on the infrastructure.

Site	Cluster	Description	N of nodes
CNR	MongoDB	MongoDB cluster is exploited in iMarine as storage for unstructured data.	4
CNR	Cassandra	The Cassandra cluster has been selected as backend for the iMarine Social Gateway functionality	2
CNR	Hadoop	Hadoop provides, among others, a distributed file system that can store data across several servers, and a platform for running tasks (Map/Reduce jobs) across those machines, running the work	12

[6] http://www.sealifebase.org/
[7] http://www.fishbase.org/
[8] http://www.i-marine.eu/Applifish/

near the data.

Table 8 - Clusters deployed on the infra

The Cassandra cluster was deployed during the reported period in order to provide the backend to the iMarine Social Gateway. It has been selected to provide horizontal scalability with the possible increase of iMarine Gateway users. In addition both MongoDB and Hadoop clusters have been enhanced in performance and storage capabilities during the second period of the project in order to reflect the increasing demand for MapReduce and NoSQL solutions.

3.2 FEDERATED RESOURCES

iMarine established close interoperability links with other infrastructures, many of them already available in the D4Science Ecosystem. This means that resources of one infrastructure can be accessed by another infrastructure and vice-versa through an agreement community-based approach under the control of the infrastructure's middleware.

The Federated Resources can be grouped in 2 categories:

- **DCI (Distributed Computing Infrastructure) Resources**
- **EA CoP Resources**

3.2.1 DCI RESOURCES

EGI

EGI s a project willing to create and maintain a pan-European Grid Infrastructure in collaboration with National Grid Initiatives (NGIs) in order to guarantee the long-term availability of a generic e-infrastructure for all European research communities and their international collaborators. The European Grid Infrastructure will (1) Operate a secure integrated production grid infrastructure that seamlessly federates resources from providers around Europe, (2) Coordinate the support of the research communities using the European infrastructure, and (3) Work with software providers within Europe and worldwide to provide high-quality innovative software solutions that deliver the capability required by user communities. EGI provides storage and computing resources, distributed across hundreds of sites worldwide and is based on different interoperable grid middleware solutions such as Globus, gLite, Arc, and Unicore. The resources offered by the EGI infrastructure significantly extend the storage and computing capacity available under the iMarine Data e-Infrastructure.

The EGI sites supporting the d4science.research-infrastructures.eu VO provide the following services to the D4Science Infrastructure:

Site Short Name	Site Official Name	CREAM	WN	SE
INFN-TRIESTE	INFN-TRIESTE	☑	☑	☑
Taiwan-LCG2	Academia Sinica Grid Computing Center	☑	☑	☑

Table 9 - EGI services supporting the d4science VO

The following table summarizes for each site the contribution to the d4science.research-infrastructures.eu VO in terms of CPU and free storage capacities.

Site	CPUs	Available Storage (TB)
INFN-TRIESTE	2380	3.7
Taiwan-LCG2	240	0,05
Total	2620	3.75

Table 10 - Availalbe EGI storage and computational resources

Compared to the previous period one site supporting our VO (Trinity College Dublin) has been dismissed while another site (INFN-BARI) decided to end the support of our VO in order to be fully dedicated to the WLCG grid. The loss of resources (especially CPUs) did not affected our operations because of the increase usage of the gCube infrastructure for computations rather than the resources coming from EGI .

VENUS-C

The VENUS-C [13] (Virtual multidisciplinary EnviroNments USing Cloud infrastructure) project was focused on developing and deploying a Cloud Computing service for research and industry communities in Europe with the aim of: (1) creating a platform that enables user applications to leverage cloud computing principles and benefits, (2) leveraging the state of the art to bring on board early adopters quickly, incrementally enable interoperability with existing Distributed Computing Infrastructures (DCIs), and push the state of the art where needed to satisfy on-boarding and interoperability, and (3) creating a sustainable infrastructure that enables cloud computing paradigms for the user communities inside the project and new communities. The VENUS-C platform is based on both commercial and open source solutions underpinned by the Engineering data centre, Microsoft through the Windows Azure and its European data centres, and two European High Performance Computing centres: The Royal Institute of Technology (KTH, Sweden) and the Barcelona Supercomputing Center (BSC, Spain).

The resources offered by VENUS-C significantly extended the storage, computing, and service hosting capacity available under the iMarine Data e-Infrastructure. These resources are accessible from the D4Science Infrastructure under a Research Collaboration Model. In particular the collaboration between iMarine and VENUS-C has granted access to Microsoft Azure [14] computation and storage cloud resources, which can be reported in the following table:

Cloud name	CPU Hours	Storage Space	End of Collaboration
MS-Azure	1.5 M	1.5 TB	April 2013

Table 11 - VENUS-C cloud details

The exploitation of the resources provided by the collaboration with VENUS-C has been concluded successfully. The usage of cloud resources (not only those provided by MS-Azure) is planned to continue during the last period of the project and will be reported within the D5.4 final deliverable.

3.2.2 EA COP RESOURCES

This section describes the set of resources close to the iMarine Ecosystem Approach Community of Practice and its scenarios that have been federated to contribute to overall set of iMarine resources:

- **Global Biodiversity Information Facility (GBIF)**[9]: this data source offers more than 377 million of records on species and more than 10,000 datasets aggregated from 400+ publishers;
- **Ocean Biogeographic Information System (OBIS)**[10]: this data source offers more that 32 million records on species and 1,000+ datasets;
- **Catalogue of Life**[11]: this data source offers an integrated checklist and a taxonomic hierarchy of more that 1.3 million species of animals, plants, fungi and micro-organisms;
- **World Register of Marine Species (WoRMS)**[12]: this data source offers species `names' for more than 200,000 species including 300,000+ species names and synonyms and 400,000+ taxa;

[9] www.gbif.org
[10] http://www.iobis.org/
[11] http://www.catalogueoflife.org/
[12] http://www.marinespecies.org/

- **Aquatic Commons**[13]: this data source offers via OAI-PMH access to thematic material covering natural marine, estuarine/brackish and fresh water environments;
- **DRS at National Institute of Oceanography**[14]: this data source, accessible via OAI-PMH, offers institutional publications including journal articles and technical reports;
- **WHOAS**[15]: this data source offers, via OAI-PMH, the production of Woods Hole scientific community including articles and data sets;
- **Central and Eastern European Marine Repository (CEEMar)**[16]: this data source offers, via OAI-PMH, material covering marine, brackish and fresh water environments;
- **OceanDocs**[17]: this data source offers, via OAI-PMH, research and publication materials in Marine Science by aggregating content form 256 repositories;
- **PANGAEA**: this data source offers, via OAI-PMH, georeferenced data from earth system research. The aggregated repositories are 475.
- **IRD UMR EME/Observatoire Thonier SDMX Registry and Repository:** This data source exposes (a) the Sardara database that contains tuna captures data from several countries, aggregated according to CWP statistical squares (1'x1' or 5'x5') and (b) the ObServe database that contains tuna and bycatches captures observed by scientific observers onboard French industrial purse seiners.

In addition of those set of resources available starting from the previous period, the following ones has been federated starting from the second period of the project:

- **The World register of Deep-Sea Species (WORDSS)[18]:** a taxonomic database of the deep-sea species based on WORMS, giving information about 20.000 species.
- **FAO Geonetwork[19]:** The FAO GeoNetwork provides Internet access to interactive maps, satellite imagery and related spatial databases maintained by FAO and its partners**.**
- **BioRisk[20]:** this data source offers, via OAI-PMH, material in ecology and biodiversity science
- **Biodiversity Heritage Library[21]:** The BHL is a consortium of natural history and botanical libraries that cooperate to digitize and make accessible the legacy literature of biodiversity held in their collections and to make that literature available for open access and responsible use as a part of a global “biodiversity commons. The data source offers via OAI-PMH access to more than 150K documents.
- **BionlineInternational:** this data source offers via OAI-PHM access to a collection of more than 30k peer reviewed bioscience journals published in developing countries**.**
- **Journal of Comparative Cryogenetics:** The Journal of Comparative Cytogenetics is a peer-reviewed, open-access, rapid online journal launched to accelerate research on all aspects of plant and animal

---

[13] http://aquaticcommons.org/
[14] http://drs.nio.org/drs
[15] http://www.mblwhoilibrary.org/services/whoas-repository-services
[16] http://www.ceemar.org/dspace/
[17] http://www.oceandocs.net/
[18] http://www.marinespecies.org/deepsea
[19] http://www.fao.org/geonetwork/srv/en/main.home
[20] http://www.pensoft.net/journals/biorisk/
[21] http://www.biodversitylibrary.org

cytogenetics, karyosystematics, and molecular systematics. The access to the documents is available through OAI-PMH.

- **Nature[22]:** this data source offers via OAI-PMH articles published in Nature.

The complete the list of OAI-PMH data sources available in the infrastructure the following smaller repositories are also accessible: International Journal of Myriapodology, Journal of Myriapodology, Journal of Hymenoptera Research, Mykokeys and Nature Conservation.

3.3 QUALITY ASSURANCE

Procedures for Quality Assurance (QA) have started to be formalized towards the end of the second period of the project (M18) in order to increase the quality of the software that is deployed on the D4Science infrastructure. As a starting point, part of the resources previously dedicated to host the Ecosystem VO and its VREs, which are no longer part of the production infrastructure, have been moved to serve the needs of the Quality Assurance activity in a the form of a pre-production infrastructure. For the sake of simplicity the VO name used to group the pre-preproduction activity has been kept the same (Ecosystem) and the same holds for the VRE that is used to validate the VRE services (TryIt).

The activity of validation of the software that is integrated by the iMarine WP7 (which could be a major, minor or a maintenance release cycle) is done at best effort by the members of WP5, and it has to be considered as an additional task to the ones included in the DoW. This means that not every gCube component released by the project is installed and analysed.

The activity has started with the components released in gCube 2.14 and it has proven to be very useful in the identification of issues that naturally cannot be discovered during the testing activities performed by WP7 members. The following points have been defined for the QA procedure:

- The infrastructure resources exploited by the QA activity are a portion of the resources allocated to the production infrastructure by the WP5 project members.
- The QA VO (Ecosystem) uses a dedicated instance of the Software Gateway pointing to a staging gCube maven repository (which could be the same as the one resulting from the integration activity)
- A separate instance of the iMarine Gateway is deployed in order to install the graphical interfaces.

The QA activities shall not last more than a week time, after this period the software installed goes into production or it is rejected because of evident issues. These points are expected to evolve and be enriched in the remaining period.

The priority is given to the following activities:

- Installation of new components not present on the production infrastructure
- Validation of the upgrade procedures (i.e. upgrade to a new version of the gCube Hosting node)
- Validation of graphical user interfaces

[22] http://www.nature.com

The part of the production resources allocated to QA task is summarized in the following table:

Partner	VMs
CNR	14
NKUA	6
Total	20

Table 12 - QA Resources

A total of 20 VMs is dedicated to this installation, which could be either used to run GHN containers, the Gateway or any additional third party software required.

4 MONITORING AND ACCOUNTING

As reported in [18], in iMarine a suite of monitoring and accounting tools is exploited:

- The monitoring and accounting of gCube resources is based on gCube services and Nagios [17]. In particular the monitoring and accounting information provided by gCube Services are based on information produced/consumed by the gCube Messaging System (MS) described in the following paragraph.
- The monitoring and accounting of UMD resources is based on tools provided by the EGI.eu project: GOCBD, SAM, MyEGI [15].
- Runtime Resources are monitored through Nagios and we are providing accounting tools only for a subset of them trough gCube Services.
- The accesses to project Gateways are then traced using the Google Analytics tool, which has been enabled starting from March 2013.

Such tools have been previously described in [18], thus only the changes with respect to the previous deliverable are being reported in this document. Therefore this part of the deliverable intends to provide statistics regarding the infrastructure load and infrastructure usage during the second period of the project. In addition statistics about the infrastructure downtimes are provided. It has to be noticed that a considerable effort has been spent during the reported period in WP8 to define a Resource Accounting Model and its tools, which will provide more fine-grained accounting data about the infrastructure exploitation to be reported on the latest deliverable of the series (D5.4).

gCube Messaging System

The gCube Messaging System has continued to be exploited to provide Accounting Information for the exploitation of the infrastructure node and Gateways. However during M18 the monitoring facilities developed in gCube have been dismissed (monitoring probes) and replaced totally by Nagios probes (see next section)

NAGIOS

The usage of NAGIOS NRPE daemons [17] on each node of the infrastructure has started during M9 and completed on M12. Such tool provides the ability to run on behalf of the central server a series of plugins and to collect the monitoring data. The following plugins have been enabled on each infrastructure node:

- nrpe-disk: checks the node disk usage
- nrpe-load: checks the load of the node
- nrpe-totalproc: checks the processes running on the node
- nrpe-swap: checks the swap space

The testing of a series of nrpe-memory plugin has also been completed. The selected plugin able to test the JVM usage of the infrastructure containers it will be installed and configured during the next project period.

In addition to local probes, the Nagios server has been equipped with a set of plugins that can remotely monitor most of the components deployed in the infrastructure. Given the usage of a well known technology (almost a defacto standard for the domain) most of the plugins installed were already developed and only in one case (the Aquamaps WebApp) a plugin has been developed by CERN. The list of services which are currently being monitored remotely by the iMarine Nagios server contains the following systems:

- GHN Containers
- Tomcat Containers
- MySQL and PostgreSQL servers
- MongoDB cluster
- Cassandra cluster
- WPS & Thredds services
- Aquamaps WebApp
- ActiveMQ MessageBroker
- Geoservers
- Nexus Maven Repository

The Nagios server and both local and remote plugins offer a powerful tool in order to notify Infrastructure and Site Managers regarding issues on the infrastructure, and it has helped to discover and promptly react critical situations.

4.1 SOME INDICATORS ON INFRASTRUCTURE USAGE

By relying on the deployed accounting tools, Figure 7 shows infrastructure logins from the iMarine and D4Science.org portals during the period M10-19. It should be noted that starting from M16 and the deployment of the gCube Social Gateway functionality the number of accesses has increased considerably, thus demonstrating the increasing interest of the users to exploit the new functionality for their daily activities. In particular the M17 login peak corresponds to presentation of the new gCube Social Gateway functionality to the iMarine Board event in March 13, which for sure has raised interest of the EA CoP participants.

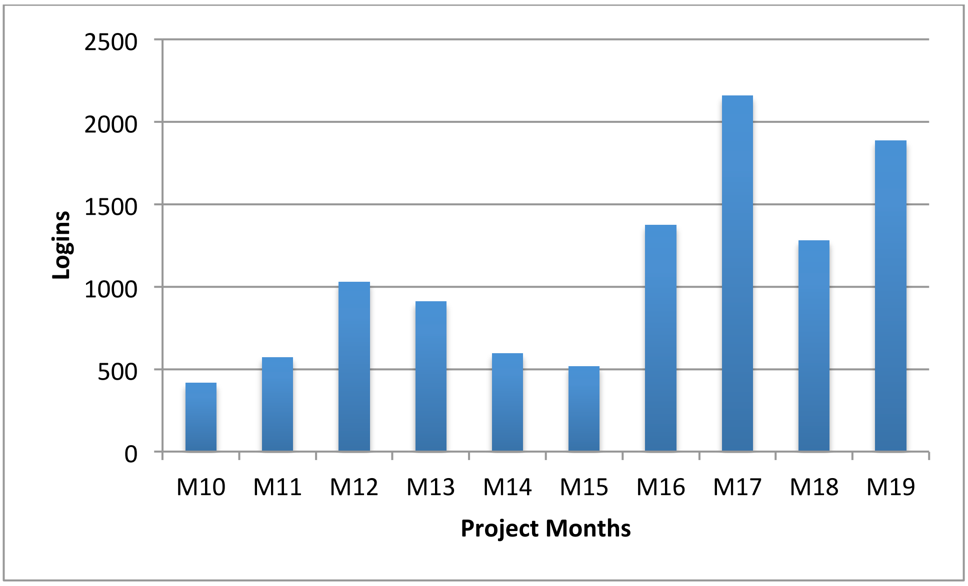

Figure 7. Infrastructure Usage: Total Number of Successful User Logins

With the deployment of the Social functionalities it was also important to understand not only the number of logins but also more detailed information about the origin, the number of pages visited etc etc. For this reason starting from March 2013 (M17), both Gateways have been configured to send statistics to Goodle

Analytics[23]. Table 13 reports the detailed statistics for the latest reported month (M19) for the iMarine Gateway. It can be noted that the top accesses comes from Europe where the iMarine project members come from (project members of course are the main users of the Gateway) but some interest is coming from other location also outside the EU.

Sub Continent Region	Visits	Pages / Visit	Avg. Visit Duration	% New Visits
	2,231 % of Total: 100.00% (2,231)	6.43 Site Avg: 6.43 (0.00%)	00:09:46 Site Avg: 00:09:46 (0.00%)	18.83% Site Avg: 18.78% (0.24%)
1. Southern Europe	1,533	6.28	00:09:09	13.44%
2. Western Europe	530	7.71	00:13:12	19.81%
3. Northern America	55	1.82	00:00:44	81.82%
4. Northern Europe	48	5.31	00:08:11	50.00%
5. South America	12	2.25	00:02:03	91.67%
6. Southern Africa	12	7.83	00:09:55	8.33%
7. South-Eastern Asia	10	4.20	00:05:39	30.00%
8. Australasia	9	5.78	00:04:16	77.78%
9. Eastern Europe	8	4.50	00:05:08	75.00%
10. Central America	5	1.20	00:00:26	80.00%

Table 13 - iMarine gateway detailed access information (M19)

Table 14 explains how much the portal users have exploited the different VOs and VREs offered by the infrastructure. This numbers however only takes into consideration the actions of the users on a subset of the functionality offered by the portal: Search, Time Series , Aquamaps , Reports and content

Operation Type	Count
Search	665
Report	372
Time Series	290
Content Retrieval	215
AquaMaps Executions	130

Table 14 - Portal Operations

As presented in Section 3.1.2 the number of gHNs running in the infrastructure at the end of M19 was in the order of one hundred. These gHNs are exploited for the deployment of other gCube services to serve particular VOs and VREs. The accounting system developed in gCube is able to gather information about the communication between gCube services. It's important to highlight that in gCube terminology we refer as a Running Instance a running service on the infrastructure and correctly registered on the Information

[23] http://www.google.com/analytics/

System. Given that, it is important to report the distribution of Running Instances as they are registered on each VOs' Information System (see Figure 8. Running Instances by VO).

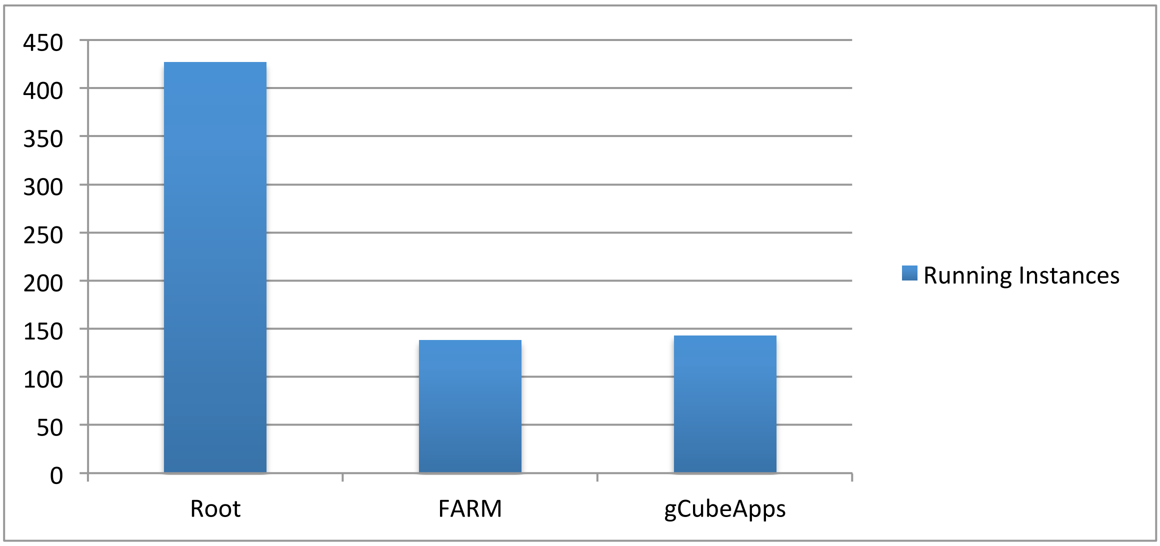

Figure 8. Running Instances by VO

The high number of Running Instances at root level can be explained by the fact that the Enabling Layer Running Instances deployed on a GHN are visible on the Information System root and this number include the Enabling Layer Running Instance of the Ecosystem VO (QA)

Due to the characteristics of the deployed monitoring tools it is possible to extract detailed information about the infrastructure load. Table 15 depicts the total number of calls accounted during the second period of the project organized by gCube subsystem.

ServiceClass	Total Invocations	Avg Invocation time (s)
InformationSystem	109,154,711	0,430
Application	70,107,708	0,018
ContentManagement	7,979,565	0,630
Index	393,584	8,210
DataAccess	204,032	5,010
Execution	118,158	0,980
DataAnalysis	74,747	0,050
VREManagement	20,962	2,450
Search	2,459	13,90
Personalisation	2,118	4,580
Messaging	1,836	10,980
DataTransfer	1,579	0,075
DataTransformation	83	7,570
OpenSearch	22	0,660

Table 15 - Infrastructure service calls statistics

The above table makes evident the large impact of the gCube Information System components in the overall infrastructure load.Some ServiceClass (Index, Search and Messaging) have an high average invocation time compared to the other classes of components, because of the nature of their synchronous blocking operations. The Application subsystem numbers include the calls to retrieve Aquamaps maps, therefore the high number of invocations testifies the interest from the community to this type of artefacts. It has to be noted that the Data Analysis invocations does not refer to the data analysis jobs duration but to the single service calls. The job duration accounting is under integration (together with other types of accounting like storage and data access) and it will be reported on the last report. Lastly a note regarding the 2 categories Content Management and DataAccess: the Content Management subsystem services have partially been replaced in production by services under the Data Access subsystems, this is why both categories are listed on the report.

More fine grained information related to single service invocations are available for Infrastructure Administrators and are used to fine tuned the deployment of each single service on the infrastructure.

4.2 INFRASTRUCTURE AVAILABILITY

As reported at the beginning of the chapter, during the second period of the project we have configured each node and service on the infrastructure to be monitored by Nagios. The server and the probes installed allow Infrastructure Manager not only to react in case of issues on the infrastructures, but also to understand the trends of availability of a category of resources, site and even a single host. For example is it possible to understand if a given site in a particular period of the year was suffering of availability issues, try to understand the reasons and take appropriate actions.

Taking in consideration the second period of the project we can report the availability trends of 2 categories of resources hosted or federated by the infra, the gCube Hosting Nodes and Runtime Resources. For the UMD resources availability the reports are generated by the different NGIs.

In the case of gCube Hosting Nodes we have an average of 99.15% of services UP (Figure 9 shows a portion of the report generated through the Nagios GUI). When compiling those numbers the portion of time where the infrastructure or part of it was in a scheduled downtime is not considered, so the percentage reported as DOWN refers mainly to network and hardware issues.

Hostgroup 'GHN' Host State Breakdowns:				
Host	% Time Up	% Time Down	% Time Unreachable	% Time Undetermined
db1.p.d4science.research-infrastructures.eu	99.148% (99.148%)	0.852% (0.852%)	0.000% (0.000%)	0.000%
dewn01.madgik.di.uoa.gr	99.155% (99.155%)	0.845% (0.845%)	0.000% (0.000%)	0.000%
dewn02.madgik.di.uoa.gr	99.157% (99.157%)	0.843% (0.843%)	0.000% (0.000%)	0.000%
dewn03.madgik.di.uoa.gr	99.154% (99.154%)	0.846% (0.846%)	0.000% (0.000%)	0.000%
dewn04.madgik.di.uoa.gr	99.157% (99.157%)	0.843% (0.843%)	0.000% (0.000%)	0.000%
dewn05.madgik.di.uoa.gr	99.138% (99.138%)	0.862% (0.862%)	0.000% (0.000%)	0.000%
dewn06.madgik.di.uoa.gr	99.135% (99.135%)	0.865% (0.865%)	0.000% (0.000%)	0.000%
dewn07.madgik.di.uoa.gr	99.157% (99.157%)	0.843% (0.843%)	0.000% (0.000%)	0.000%
dewn08.madgik.di.uoa.gr	99.157% (99.157%)	0.843% (0.843%)	0.000% (0.000%)	0.000%
dewn09.madgik.di.uoa.gr	99.160% (99.160%)	0.840% (0.840%)	0.000% (0.000%)	0.000%
dewn10.madgik.di.uoa.gr	99.161% (99.161%)	0.839% (0.839%)	0.000% (0.000%)	0.000%
dl16.di.uoa.gr	99.130% (99.130%)	0.870% (0.870%)	0.000% (0.000%)	0.000%
dl17.di.uoa.gr	99.131% (99.131%)	0.869% (0.869%)	0.000% (0.000%)	0.000%
dl18.di.uoa.gr	99.132% (99.132%)	0.868% (0.868%)	0.000% (0.000%)	0.000%
dl19.di.uoa.gr	99.095% (99.095%)	0.905% (0.905%)	0.000% (0.000%)	0.000%
dl20.di.uoa.gr	99.036% (99.036%)	0.964% (0.964%)	0.000% (0.000%)	0.000%
dl21.di.uoa.gr	99.069% (99.069%)	0.931% (0.931%)	0.000% (0.000%)	0.000%
dl23.di.uoa.gr	99.121% (99.121%)	0.879% (0.879%)	0.000% (0.000%)	0.000%

Figure 9 - Availability Report for GHNs

Looking at the availability report for the gCube Runtime Resources category (see Figure 10), the average for the second period of the project is also quite satisfying (99,31%), even in the case of federated resources that are not under our administration.

Hostgroup 'gCube Runtime Resource' Host State Breakdowns:				
Host	% Time Up	% Time Down	% Time Unreachable	% Time Undetermined
data.d4science.org	99.997% (99.997%)	0.003% (0.003%)	0.000% (0.000%)	0.000%
dev.d4science.org	99.997% (99.997%)	0.003% (0.003%)	0.000% (0.000%)	0.000%
geoserver.d4science-ii.research-infrastructures.eu	99.245% (99.245%)	0.755% (0.755%)	0.000% (0.000%)	0.000%
geoserver2.d4science.research-infrastructures.eu	99.280% (99.280%)	0.720% (0.720%)	0.000% (0.000%)	0.000%
geoserver3.d4science.research-infrastructures.eu	99.233% (99.233%)	0.767% (0.767%)	0.000% (0.000%)	0.000%
geoserver4.d4science.research-infrastructures.eu	99.272% (99.272%)	0.728% (0.728%)	0.000% (0.000%)	0.000%
maven.research-infrastructures.eu	99.212% (99.212%)	0.788% (0.788%)	0.000% (0.000%)	0.000%
message-broker.d4science.research-infrastructures.eu	95.573% (95.573%)	4.427% (4.427%)	0.000% (0.000%)	0.000%
node1.p.cassandra.research-infrastructures.eu	99.969% (99.969%)	0.031% (0.031%)	0.000% (0.000%)	0.000%
node2.p.cassandra.research-infrastructures.eu	99.916% (99.916%)	0.084% (0.084%)	0.000% (0.000%)	0.000%
node58.p.d4science.research-infrastructures.eu	99.150% (99.150%)	0.850% (0.850%)	0.000% (0.000%)	0.000%
node67.p.d4science.research-infrastructures.eu	99.148% (99.148%)	0.852% (0.852%)	0.000% (0.000%)	0.000%
node73.p.d4science.research-infrastructures.eu	99.264% (99.264%)	0.736% (0.736%)	0.000% (0.000%)	0.000%
node76.p.d4science.research-infrastructures.eu	99.903% (99.903%)	0.097% (0.097%)	0.000% (0.000%)	0.000%
node80.p.d4science.research-infrastructures.eu	99.185% (99.185%)	0.815% (0.815%)	0.000% (0.000%)	0.000%
thredds.research-infrastructures.eu	99.999% (99.999%)	0.001% (0.001%)	0.000% (0.000%)	0.000%
wps01.i-marine.d4science.org	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
Average	99.314% (99.314%)	0.686% (0.686%)	0.000% (0.000%)	0.000%

Figure 10 - Availability Report for gCube Runtime Resources

4.2.1 INFRASTRUCTURE DOWNTIMES

Figure 11 presents the total number of downtimes declared by the sites that host the iMarine resources of the D4Science infrastructure.

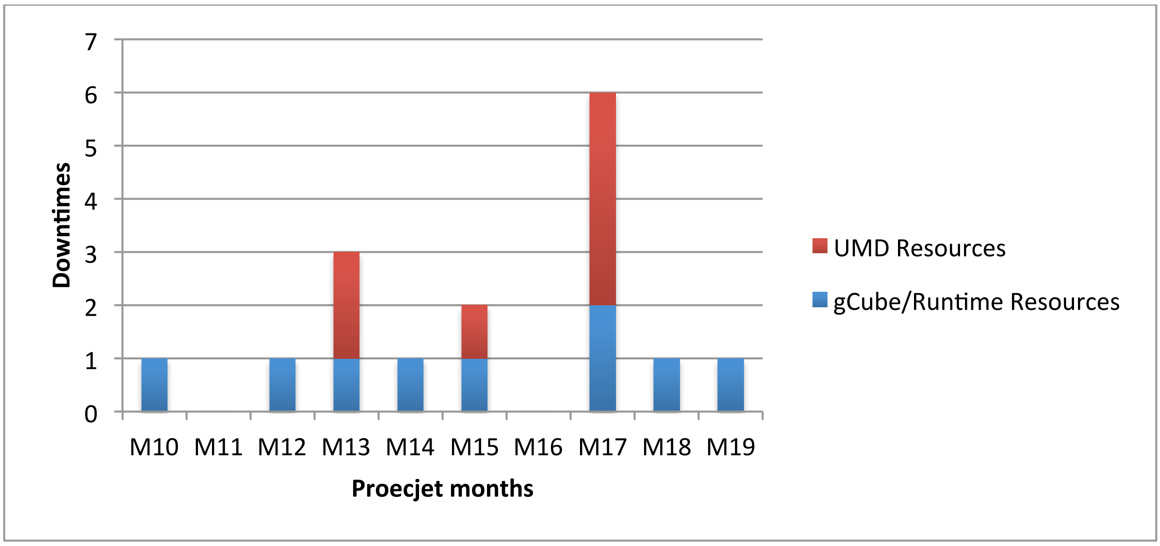

Figure 11. Infrastructure Resources Downtimes

It should be mentioned that these "downtimes" were caused either by scheduled network interventions or by infrastructure upgrades. The downtimes lasted in average 4 hours (infrastructure upgrades took also 8 hours), so the production infrastructure was never affected by long unavailability periods and each single VRE was unavailable for just a short period of time. In downtimes triggered by infrastructure upgrades gCube downtimes and Runtime Resources downtimes are combined since usually the upgrades to both resources are performed at the same time. Moreover, the high number of UMD downtimes on M17, can be explained by the upgrade of both our grid sites to the new middleware (from UMD to UMD2).

5 PRODUCTION SUPPORT

This section describes the activities carried out to provide support to the operation and exploitation of the D4Science infrastructure by its different user types: VO Administrators, VRE Managers, Site Managers, etc. The infrastructure support activity is based on the incident management procedure [16]. This procedure follows the ITIL methodology for incident management and has been adopted since the beginning of the project and it has been enhanced during the project lifetime.

The incident management procedure description is available on the D5.1 deliverable [11] together with detailed information about the procedures, the people involved, and the workflow of the different steps.

Even if there is a common incident management procedure, some small differences apply whether the procedure is applied to gCube, Runtime Resource or UMD nodes.

- gCube and Runtime Resources: All incidents related to the exploitation and deployment of gCube and Runtime Resources are tracked according to the incident management procedure;
- UMD Nodes: The incidents related to the usage of UMD nodes are followed according to the incident management procedure. The only exception concerns the incidents related to the deployment of UMD nodes, which are submitted directly to the regional NGIs.

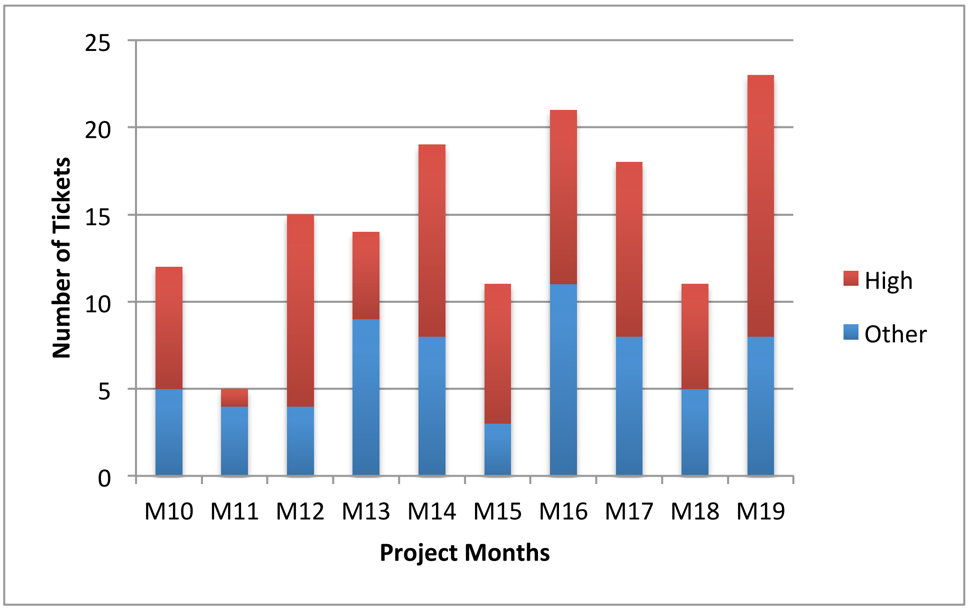

Figure 12. Incident Tickets by Priority

A total number of 149 tickets were submitted and closed during the second period of the project. From this total, 56.37% (84 tickets) was high priority incidents. Figure 12 gives detailed information about the total number of submitted tickets highlighting as well the number of high priority tickets.

Concerning the affected environments, Figure 13 shows the distribution of the recorded incident tickets across the support VREs. Many tickets were common to all existing VREs. Looking to VRE-specific tickets the most affected VREs are the ones under the FARM VO (FCPPS, ICIS, and AquaMaps) and the BiodiversityResearchEndironement VRE, which also result to be the more, exploited. Please note that the figure does not report incidents not specific to a particular environment (97 tickets)

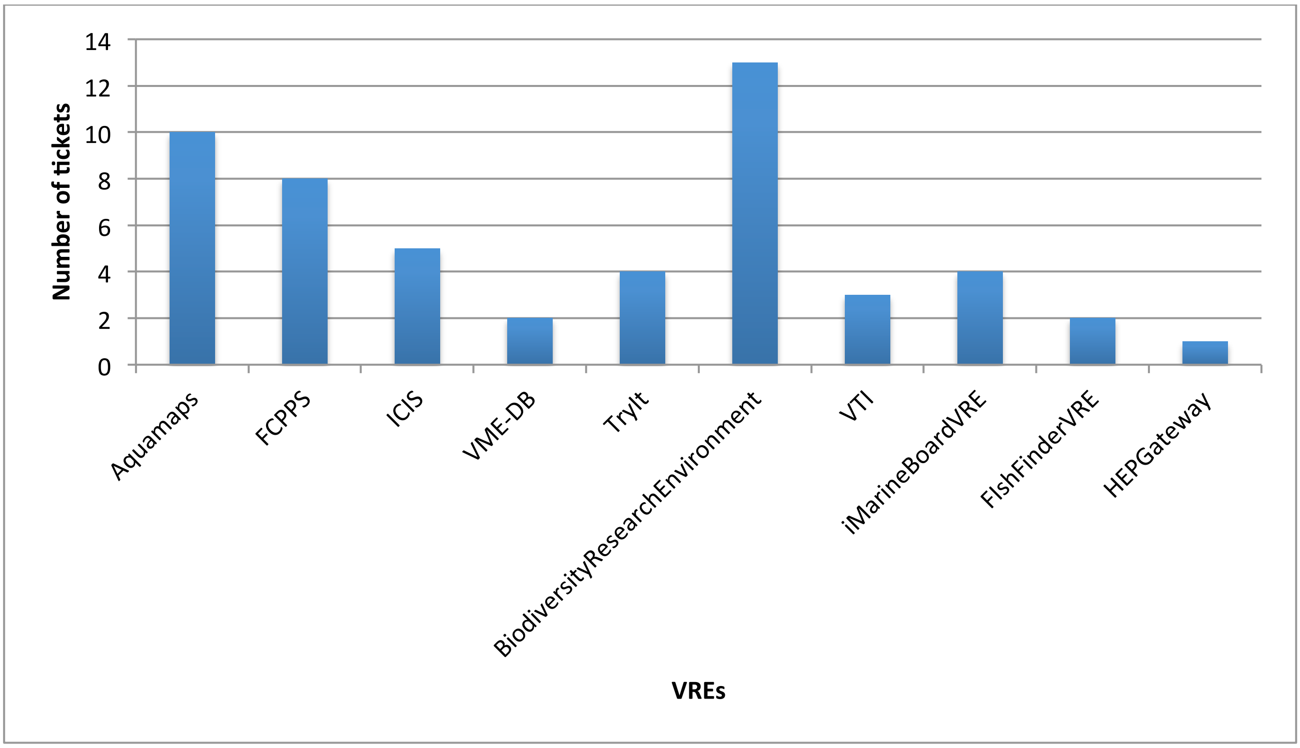

Figure 13 - Incident Tickets by VRE

REFERENCES

[1] D4Science.org Web site http://www.d4science.org/
[2] iMarine Web site http://www.i-marine.eu/
[3] EuBrazilOpenBio Web Site http://www.eubrazilopenbio.eu
[4] ENVRI Web site http://envri.eu/
[5] Ecosystem Approach Community of Practice Wiki page http://wiki.i-marine.eu/index.php/Ecosystem_Approach_Community_of_Practice_Overview
[6] iMarine. Ecosystem Approach Community of Practice: Requirements http://wiki.i-marine.eu/index.php/Ecosystem_Approach_Community_of_Practice:_Requirements
[7] L. Candela. Virtual Research Environments Plan. D6.1 iMarine Project Deliverable. January 2012
[8] The UMD Distribution Web Site http://repository.egi.eu/category/umd_releases/
[9] The EMI Project Web Site http://www.eu-emi.eu/
[10] The VOMS server admin page https://voms.research-infrastructures.eu:8443/voms/d4science.research-infrastructures.eu
[11] A. Manzi. iMarine Data e-Infrastructure Plan. D5.1 iMarine Project Deliverable. January 2012
[12] The EGI Infrastructure Web site http://www.egi.eu/
[13] The VENUS-C project Web Site http://www.venus-c.eu
[14] The MS Azure Web Site http://www.windowsazure.com/
[15] The MyEGI Web Site https://tomtools.cern.ch/confluence/display/SAM/MyEGI
[16] The iMarine incident management procedure http://wiki.i-marine.eu/index.php/Procedure_Infrastructure_Incident_Management
[17] The Nagios Web Site http://www.nagios.org/
[18] A. Manzi. L. Candela iMarine Data e-Infrastructure Operation Report. D5.2 iMarine Project Deliverable. Sept 2012