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.	**D5.4**
Delivery Date	***September 2014***
Author	***Leonardo Candela, Roberto Cirillo, Andrea Manzi, Pasquale Pagano***

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 third of the series of reports on the activities performed while operating such a system. The deliverable describes the state of the Data e-Infrastructure in terms of nodes available, software deployed, quality of the service, and usage as of September 2014.*

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.	D5.4
Deliverable Title	iMarine Data e-Infrastructure Operation Report
Contractual Delivery Date	September 2014
Actual Delivery Date	September 2014
Author(s)	Leonardo Candela (CNR), Roberto Cirillo (CNR), Andrea Manzi (CERN), Pasquale Pagano (CNR)
Editor(s)	Leonardo Candela (CNR)
Reviewer(s)	George Kakaletris (NKUA)
Contributor(s)	Kostantinos Kakaletris (NKUA), Valentina Marioli (CNR), Tommaso Piccioli (CNR)
Work Package No.	WP 5
Work Package Title	iMarine Data E-Infrastructure Deployment and Operation
Work Package Leader	Andrea Manzi (CERN)
Work Package Participants	CERN, CNR, NKUA, FAO, FIN
Estimated Person Months	2
Distribution	Public
Nature	Report
Version / Revision	1.0
Draft / Final	Final
Total No. Pages (including cover)	36
Keywords	Infrastructure Operation; Federated Resources; Virtual Organisation;

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

Community of Practice (CoP): 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: The Disk Pool Manager is a lightweight solution for disk storage management. If offers the required SRM interfaces, hopefully without being complicated by other modes of access or complications such as tape storage systems. It has been developed at CERN.

EAF: Ecosystem Approach to Fisheries. This approach strives to balance diverse societal objectives, by taking into account the knowledge and uncertainties about biotic, abiotic and human components of ecosystems and their interactions and applying an integrated approach to fisheries within ecologically meaningful boundaries.

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.

LFC: LCG File Catalog. It is a File Catalog, i.e. a service mapping logical file names to storage URL(s), developed by the LCG project.

LMEs: Large Marine Ecosystems. LMEs are regions of the world's oceans, encompassing coastal areas from river basins and estuaries to the seaward boundaries of continental shelves and the outer margins of the major ocean current systems.

NGI: National Grid Initiative. NGIs are organisations set up by individual countries to manage the computing resources they provide.

QA: Quality Assurance.

sBDII: Site BDII. Every Site BDII hosts a BDII, i.e. an information system for Grid Computing infrastructures. It consists of a standard LDAP server which is updated by an external process.

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.

Virtual Organisation (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.

VOMS: The Virtual Organisation Membership Service is a system for managing a database of user accounts. It was initially developed in the context of the DataGrid project.

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 [3]. Such an ecosystem is a result from the joint effort and resources of a number of projects including iMarine [6], EUBrazilOpenBio [4], and ENVRI [5]. 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 [7].

This report documents the work done to operate the iMarine resources part of the D4Science infrastructure during the period from June 2013 to September 2014, which corresponds to months M20 to M35 of the iMarine project.

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 ***20 VREs*** were provided. Overall, these VOs and VREs have served more than 750 users. In additiont to that, more than 400 additional users have been served via other portals (e.g. EGIP) and VREs (e.g. ENVRI) developed and operated by relying on gCube and iMarine technologies and procedures.

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.

Through the iMarine portal a total of ***14,419 sessions*** were served with an average of approximately 900 sessions per month. These sessions were started by a total of ***2,639 users*** with an average of 221 users per month.

A straightforward support procedure continued to be implemented during the course of this project months. A total number of ***237 tickets*** were submitted and closed during the reporting period. The 53% (126 tickets) of the managed tickets was high priority incidents. The 70% (166 tickets) of tickets were not specific to any VRE, while the 10% (25 tickets) of tickets was reported by referring to FARM VREs and the 19% (46) of tickets was reported by referring to gCubeApps VREs.

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 (d) presents the activity performed by the production support team to deal with issues that emerged 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 third period of the iMarine project to operate such resources. Before describing the resources which compose the infrastructure, it's important to highlight the main aspects that characterize 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 of 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 [11]. 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 [12]. 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 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.

In the previous deliverable of the series D5.3 [10], the aforementioned concepts are described in detail.

This deliverable builds on D5.3 and updates that document to describe the state of the art at September 2014. During the third period of the project the focus was on consolidating and enhancing the resources needed by iMarine VOs and VREs.

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

These requirements are documented in "Ecosystem Approach Community of Practice: Requirements" [8]. Specified requirements are carefully analysed and transformed into a development plan as documented in "Virtual Research Environment Deployment and Operation: Plan" [1]. 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 M35, i.e. September 2014. A detailed report is in D6.5 [2].

2.1 VIRTUAL ORGANISATIONS (VOS)

During this reporting period two (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.

Table 1. iMarine Virtual Organisations

VO	VREs	Users[2]
FARM	10	278
gCubeApps	10	492

[1] The D4Science and D4Science II projects User Communities

[2] The number of users of a VO is the sum of the number of users of the VREs operated in the context of that VO.

2.2 VIRTUAL RESEARCH ENVIRONMENT (VRES)

Twenty (20) 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 ten (10) 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 **MarineSearch** Virtual Research Environment is for providing the members of the EA-CoP with an environment dedicated to showcase the data discovery facilities, in particular the IR and semantic based ones. The information space, is tailored to marine related data sources.

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 **Scalable Data Mining** Virtual Research Environment is designed to apply data analytics algorithms to biological data. Algorithms include niche modelling, Bayesian models, trends discovery.

The **Too Big To Ignore (TBTI)** Virtual Research Environment is for the members of the Too Big to Ignore[3] initiative. In particular, the VRE is oriented to provide the members of this community with an environment demonstrating how the iMarine set of facilities can be used to serve the needs of this community that is

[3] http://toobigtoignore.net Too Big To Ignore Initiative website

established to rectify the marginalization of small-scale fisheries in national and international policies, and to develop research and governance capacity to address global fisheries challenges.

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).

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.

2.2.2 THE GCUBEAPPS VRES

The gCubeApps Virtual Organisation hosts and operates the following ten (10) Virtual Research Environments.

The **BiodiversityLab** Virtual Research Environment is for scholars willing to perform complete experiments about single individuals or groups of marine species.

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.[4]

The **BiOnym** Virtual Research Environment is for taxonomists and biodiversity scientists dealing with species taxonomic information. The main facility is BiOnym, an application specifically conceived to compare species scientific names against reference lists.

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 **iSearch** Virtual Research Environment is for providing its users with an environment dedicated to showcase the data discovery facilities, in particular the IR and semantic based ones.

The **TabularDataLab** Virtual Research Environment is for providing data managers with a feature rich environment for supporting the management of tabular data, i.e. any dataset that can be represented in a table format, and code lists. In practice, this environment is aiming at replacing and reinforcing the facilities offered by the TimeSeries VRE with new services.

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.

[4] This VRE was dismissed during September 2014 since its facilities largely overlap with the BiodiversityLab VRE. Its users have been automatically subscribed to the BiodiversityLab VRE.

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

Table 2 reports some indicators on Virtual Research Environments operated during the reporting period, namely the creation date and the number of users.

Table 2. iMarine Virtual Research Environments detailed information

VRE	VO	Creation date[5]	Users[6]
AquaMaps	FARM	M1, Nov 2011	57
BiodiversityLab	gCubeApps	M26, Dec 2013	118
BiodiversityResearchEnvironment	gCubeApps	M9, July 2012	41
BiOnym	gCubeApps	M30, Apr 2014	16
DocumentsWorkflow	gCubeApps	M1, Nov 2011	44
EcologicalModeling	gCubeApps	M1, Nov 2011	73
FCPPS	FARM	M1, Nov 2011	35
FishFinderVRE	FARM	M16, Mar 2013	15
ICIS	FARM	M1, Nov 2011	44
iSearch	gCubeApps	M21, Aug 2013	35
iMarineBoardVRE	FARM	M15, Feb 2013	28
MarineSearch	FARM	M29, Mar 2014	16
ScalableDataMining	FARM	M17, Oct 2012	27
TabularDataLab	gCubeApps	M32, Jun 2014	17
TBTI	FARM	M20, Jul 2013	11
TCom	gCubeApps	M17, Apr 2013	41
TimeSeries	gCubeApps	M1, Nov 2011	60
VesselActivitiesAnalyzer	gCubeApps	M1, Nov 2011	47
VME-DB	FARM	M9, Jul 2012	17
VTI	FARM	M1, Nov 2011	28

The following three figures describe the evolution in term of users served by a given VRE with respect to the previous reporting period [10]. 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.

[5] With M1 we highlight the fact that this VRE was created in the context of previous projects and maintained by iMarine.
[6] At the date of writing this report, i.e. September 2014.

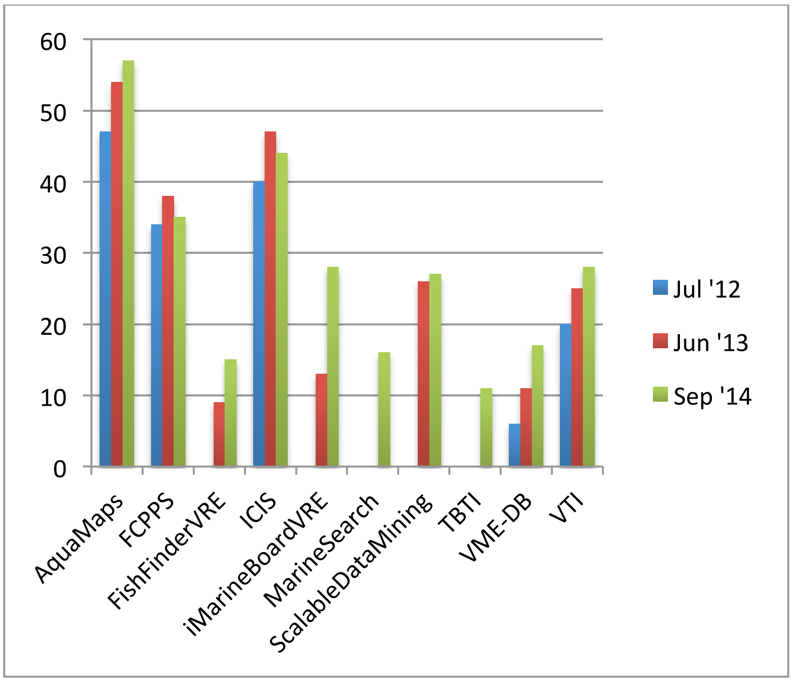

Figure 1. FARM VO: Evolution of number of users per VRE

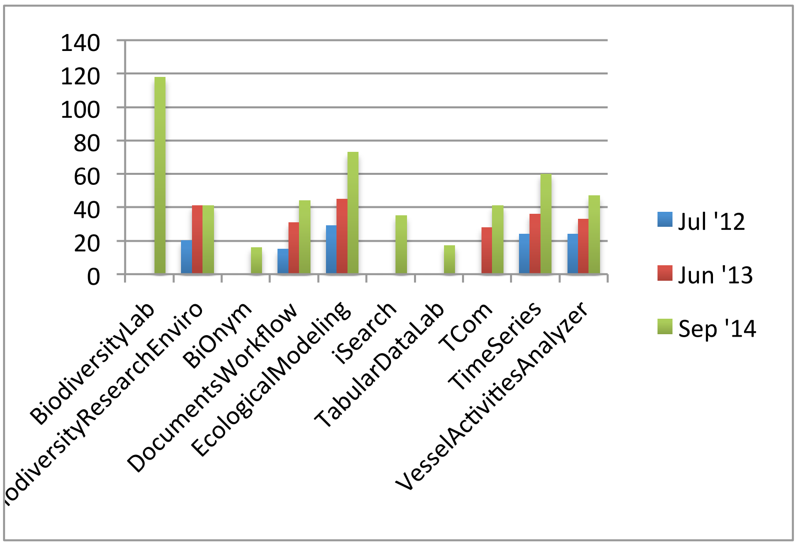

Figure 2. gCubeApps VO: Evolution of number of users per VRE

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 +24% (223 at Jun '13 vs 278 at Sep '14) in the number of users served by FARM VREs and of +129% (214 at Jun '13 vs 492 at Sep '14) in the number of users served by gCubeApps VREs.

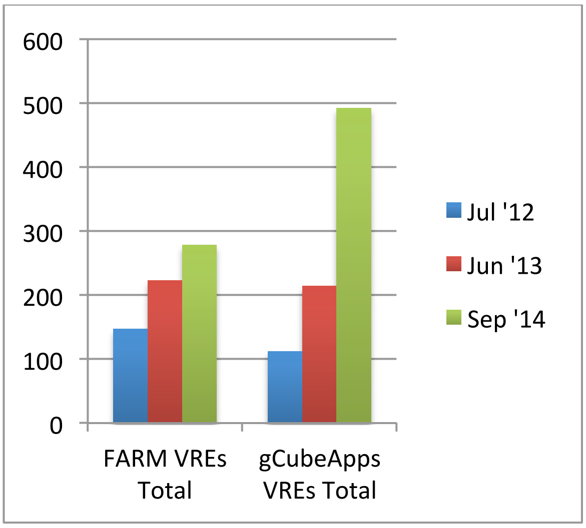

Figure 3. Evolution of total number of users served by VREs in a given VO

2.3 GATEWAYS

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

- **iMarine gateway**[7]: it provides access to the FARM VO and its VREs and the gCubeApps VO and its VREs;
- **D4science.org gateway** [https://portal.d4science.org[8]: it provides access to the the gCubeApps VO and its VREs.

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).

The iMarine gateway is dedicated to serve the iMarine community only. The figures reported in the previous sections as well as in Section 4.1.1 for the number of users refers to this gateway.

The D4Science gateway is actually serving a larger community. Because of this, it is offering the gCubeApps Virtual Research Environments as well as other Virtual Research Environments like the ENVRI VRE. The ENVRI VRE was developed in the context of the homonymous project with the aim to provide the members of this community with a working environment based on gCube technologies. It is currently serving 24 selected users spread across diverse Institutions and research centres.

Besides the above mentioned portals, two additional scenarios have been served by relying on the technology and procedures developed by the iMarine project, i.e. Social-ISTI and EGIP.

Social-ISTI[9] is a gCube based solution for supporting the communication among the members of the Istituto di Scienza e Tecnologie dell'Informazione of the National Research Council of Italy. It is currently serving 340 users.

[7] IMarine Gateway https://portal.i-marine.d4science.org/
[8] D4Science Gateway https://portal.d4science.org/
[9] Social-ISTI web site https://social.isti.cnr.it

EGIP[10] (European Geothermal Information Platform) is a gCube based solution for providing the members of the European Geothermal community (i.e., scientific, political and industrial stakeholders) – aggregated via the Geo ERA-NET project – with the nucleus of facilities and data envisaged by the fully fledged EGIP use case. It contains the most urgent information and some main functionality to demonstrate the effectiveness and efficiency of the envisaged platform. It is currently serving 24 selected users spread across diverse Institutions and research centres.

[10] EGIP web site https://egip.d4science.org

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. Sec. 3.1), i.e. resources deployed by the iMarine project partners and thus owned by the D4Science infrastructure;
- Federated Resources (cf. Sec. 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. Sec. 3.1.2.1) in order to maintain the infrastructure operational and the activities and resources dedicated to Quality & Assurance tasks (cf. Sec. 3.3).

3.1 HOSTED RESOURCES

As originally defined in deliverable D5.1 iMarine Data e-Infrastructure Plan [9], the D4Science infrastructure hosting resources dedicated to iMarine are provided by T5.2 tasks members plus an external partner (ASGC Taiwan). Moreover, by continuing the offering started in the previous period, the VLIZ project partner contributed to the D4Science 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 also provided resources, i.e. **ASGC** – Taiwan.

3.1.1 HOSTING RESOURCES

Table 3 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** (HW) together with type of CPU or **Virtual Machines** (VM) and the type of virtualization system.

Table 3. Hosting Resources by Site

Site	Type	RAM (GB)	Disk (TB)	CPUs
CNR	VM / Xen hypervisor	1228	44	834
NKUA	VM / Xen hypervisor	114	1.2	63
ASGC	HW / Two Quad-Core Intel(R) Xeon(R) CPU 5130 @ 2.00GHz	8	0.2	8
	HW / Quad-Core Intel(R) Xeon(R) CPU 5150 @ 2.66GHz	4	0.1	4
FAO	HW / Two Quad-Core Intel Xeon X5450 @ 3.0 GHz	8	0.3	8
VLIZ	HW / Intel(R) Xeon(R) CPU E5649 @ 2.53GHz	3	0.4	4
	Total	**1366**	**46.9**	**918**

The hosting resources deployed are compliant with the plans defined at the beginning of the project and allowed the deployment and availability of all VOs and VREs requested 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)[11]. 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.

Table 4 presents the number of gCube hosting nodes deployed per site.

Table 4. gCube Hosting Node per site

Partner	gCube Hosting Nodes
CNR	61
NKUA	22
ASGC	3
FAO	2
VLIZ	1
Total	**89**

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.

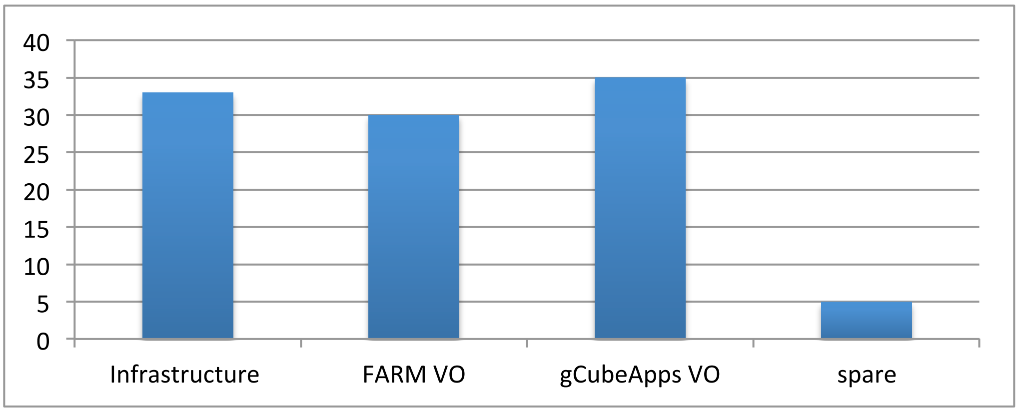

Figure 4. gCube Hosting Node per VO

3.1.2.1 GCUBE DEPLOYMENT

The latest gCube release deployed in production as of September 2014 was the release gCube 3.4.0, deployed at the end of the September '14. During the reporting period (June '13 – September '14), the infrastructure operation was based on a total of 11 gCube releases, namely gCube 2.15.0 (Jun '13), gCube

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

2.16.0 (Jul ‘13), gCube 2.16.1 (Sep ’13), gCube 2.17.0 (Dec ’13), gCube 2.17.1 (Dec ‘13), gCube 3.0.0 (Feb ’14), gCube 3.1.0 (May ’14), gCube 3.1.1 (Jun ’14), gCube 3.2.0 (Jun ’14), gCube 3.3.0 (Jul ’14), and gCube 3.4.0 (Sep ‘14).

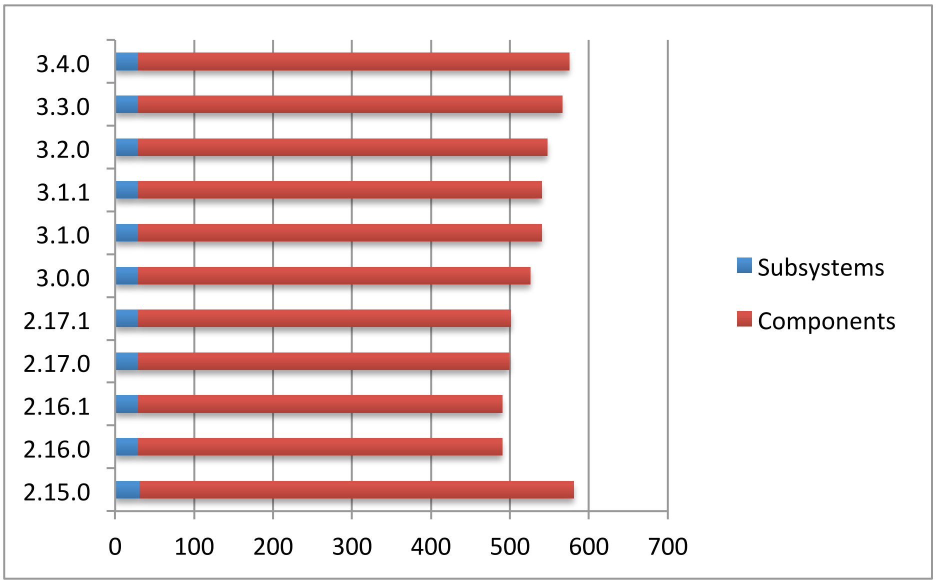

Figure 5. gCube Releases Size

The decrease of the number of components contributing to form gCube 2.16.0 with respect to the previous release, i.e. gCube 2.15.0, was due to the dismissal of two subsystems and the related components. In particular, the ‘ontology-management’ and ‘dir’ subsystems, two very old subsystems that have been supplanted by others from the functional perspective, was dismissed. From gCube 2.16.0 on the number of subsystems remained stable to 29 components while the number of components contributing to form these subsystems growth up to the 546 of the gCube 3.4.0 release.

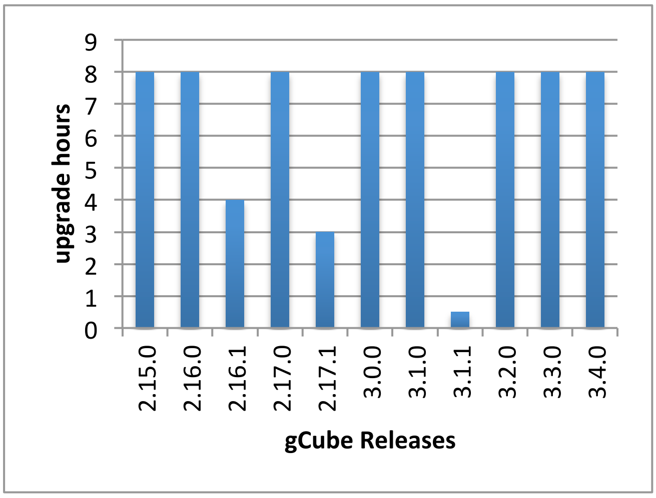

Figure 6. gCube Releases deployment time

Figure 6 shows the number of hours needed to deploy any new gCube release in the infrastructure. The deployment of every major or minor release was scheduled and effectively required 8 hours while the deployment of a maintenance release required a minimum on half an hour (gCube 3.1.1) to a maximum of 4 hours (gCube 2.16.1). It should be noted that usually the upgrade resulting from a new major or minor release required modifications to all infrastructure nodes.

3.1.3 UMD RESOURCES

Two sites (out of the four project sites) offered UMD resources to the infrastructure. UMD resources are computing and storage services running UMD middleware [11], which is a distribution for EGI of the software developed by the EMI project [12]. The UMD services are exploited by gCube services, which then provide higher-level functionality through the iMarine VREs. Moreover, all UMD services provided by iMarine partners are also registered to join the EGI production infrastructure (cf. Section 3.2.1).

During the third period of the project, the two sites previously maintained and described in Table 5 have been operated up to April 2014. UMD resources decommission was decided because the exploitation of such resources was considered no longer fundamental with respect to the iMarine project operation needs.

Table 5. UMD services per site

	CREAM CE	WN	sBDII	SE	WMS	VOMS	LFC	APEL	UI	LB
CNR	✓	✓	✓	✓	✓	✓	✓		✓	✓
NKUA	✓	✓	✓	✓	✓			✓	✓	✓

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 the following categories: third party services, Database, and clusters.

Third party services are software systems and technologies not belonging to the gCube technology yet supporting the operation of it. They contribute to form the infrastructure and are made available to gCube components. They include *repositories for various data types*, e.g. JackRabbit, THREDDS, and Geoserver, *Catalogues*, e.g. SDMX Registry, GeoNetwork, and supporting services, e.g. ActiveMQ, RStudio. A detailed list is in Table 6.

Table 6. Third party services hosted by the D4Science infrastructure

Site	Service	Description	Number of Instances
CNR	JackRabbit Repository	A JackRabbit repository is used as a backend for User workspace data.	1
	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
	GeoServer	The GeoServer service is deployed in the D4science infrastructure to enable the building of an SDI. The Service is deployed together with an instance of PostgreSQL DB, in order to store GIS data.	4
	GeoNetwork	The GeoNetwork service is deployed in the D4Science infrastructure to enable the building of an SDI. It is a catalogue federating other GeoNetworks as well as the content stored in the GeoServer and THREDDS instances.	1
	THREDDS	The Thredds Service is responsible for the publication and discovery of geospatial data by means of OGC	1

		standard protocols. It is another repository for geospatial data contributing to build the SDI.	
	North52 WPS	The North52 WPS service implements standard geospatial processing interface. It offers a number of algorithms.	1
	SDMX Registry	The SDMX Registry allows the publication and registration of SDMX Data Sources.	1
	RStudio	The RStudio service is offering an environment for the execution of R scripts.	1
	Nexus Server	Nexus hosts the Maven Repositories containing gCube artefacts	1
Terradue	North52 WPS	The North52 WPS service implements standard geospatial processing interface. It has been integrated in iMarine with the Hadoop backend in order to increase the service scalability. It offers a number of algorithms.	1
IRD	North52 WPS	The North52 WPS service implements standard geospatial processing interface. This instance offers a number of algorithms developed by IRD.	1

Databases represent another typology of runtime resource contributing to form the D4Science infrastructure. These databases are either used to form the persistence backend of some services, e.g. Tabular Data Manager and Aquamaps, or to host datasets of interest for many services, e.g. FishBase and SeaLifeBase. A detailed list is in Table 7.

Table 7. Databases hosted by the D4Science Infrastructure

Site	Database	Description	Number of Instances
CNR	PostgreSQL	The PostgreSQL DB backend for the gCube TimeSeries and Aquamaps Services	4
	PostgreSQL	The PostgreSQL DB containing conversion tables between Fish Codes	1
	PostgreSQL	The PostgreSQL DB backend hosting a mirror of the FishBase and SeaLifeBase DBs	1
	PostgreSQL	The PostgreSQL DB hosting the OBIS Biodiversity Database	1
	PostgreSQL	The PostgreSQL DB backend for the gCube Statistical Manager service	1
	PostgreSQL	The PostgreSQL DB backend for the gCube Tabular Data Manager service	1
	PostgreSQL	The PostgreSQL DB containing geospatial data for GeoServer and GeoNetwork	1
	MySQL	The MySQL DB hosting the ICTV biodiversity repository	1
	MySQL	The MySQL DB storing the D4Science Infrastructure Accounting and Monitoring data	1
NKUA	PostgreSQL	The PostgresSQL DB hosting the FLORA biodiversity repository	1
VLIZ	PostgreSQL	The PostgreSQL DB hosting the OBIS biodiversity Database	1

The third typology of runtime resource is represented by clusters. These resources are 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. **Table 8** gives details on the cluster deployed in the infrastructure.

Table 8. CLusters deployed in the D4Science infrastructure

Site	Cluster	Description	Number of Nodes
CNR	MongoDB	MongoDB cluster is exploited in iMarine as storage for unstructured data.	4
	Cassandra	The Cassandra cluster has been selected as backend for the iMarine Social Gateway functionality.	2
	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 near the data.	12

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 (cf. Sec. 3.2.1).
- EA CoP Resources (cf. Sec. 3.2.2).

3.2.1 DCI RESOURCES

In the case of Computing Infrastructures, the infrastructure was configured to interface with EGI and VENUS-C. Details have been reported in the previous deliverable [10].

In the third reporting period it was decided to remove the links with these initiative since they were considered no longer fundamental to support the user communities and the relative use cases target of the iMarine project. The loss of resources (especially CPUs) did not affected the operation of the infrastructure because of the increase usage of the gCube infrastructure for computations rather than the resources coming from EGI.

Overall, this decommission does not affected at all the operation of the iMarine infrastructure. The enabling technology proved to be able to interface with EGI as well as to exploit Microsoft Azure-based infrastructures. Thus when needed to enlarge the computing capacity of the infrastructure it will be sufficient to properly configure theenabling technology and to establish collaboration practices with the providers.

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. In particular, these resources represent databases and information systems giving access to datasets of interest for the Community. These databases and information systems include species data (cf. Sec. 3.2.2.1), geospatial data (cf. Sec. 3.2.2.2), statistical data (cf. Sec. 3.2.2.3) and other research data including papers (cf. Sec. 3.2.2.4).

3.2.2.1 SPECIES DATA

The following databases and information systems have been integrated to give access to species data, both taxonomic data and occurrence data:

- Catalogue of Life: The data source offers an integrated checklist and a taxonomic hierarchy of more than 1.3 million species of animals, plants, fungi and micro-organisms.
- FAO ASFIS: The List of Species for Fishery Statistics Purpose includes 12,000+ species of interest or relations to fisheries and aquaculture; www.fao.org/fishery/collection/asfis/en
- GBIF: The data source offers more than 430 million of records on species and more than 14,000 datasets aggregated from 580+ publishers; www.gbif.org
- Fishbase: The data source offers access to 32,700 Species, 302,900 Common names, 53,600 Pictures, 49,700 References aggregated thanks to the effort of thousand collaborators.
- IRMNG: The Interim Register of Marine and Nonmarine Genera data source offers access to over 465,000 genus names and 1.6 million species names; www.obis.org.au/irmng
- ITIS: The Integrated Taxonomic Information System data source offers authoritative taxonomic information on plants, animals, fungi, and microbes of North America and the world; http://www.itis.gov
- NCBI Taxonomy: The National Center of Biotechnology Information data source offers a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet; www.ncbi.nlm.nih.gov/taxonomy
- OBIS: The Ocean Biogeographic Information System data source offers more that 40 million records on species and 1,600+ datasets; www.iobis.org
- SeaLifeBase: The data source offers access to 126,000 Species, 27,300 Common names, 11,900 Pictures, 18,200 References aggregated thanks to the effort of hundred collaborators.
- WoRMS: The World Register of Marine Species data source offers species "names" for more than 200,000 species including 300,000+ species names and synonyms and 400,000+ taxa; http://www.marinespecies.org
- WoRDSS & The World Register of Deep-Sea Species data source offers species "names" for deep-sea species based on WoRMS. http://www.marinespecies.org/deepsea

3.2.2.2 GEOSPATIAL DATA

The following databases and information systems have been integrated to give access to geospatial data:

- FAO GeoNetwork: The data source exposes spatial data maintained by FAO and its partners; http://www.fao.org/geonetwork
- World Ocean Atlas: The data source gives access to a number of environmental variables. In particular, iMarine focuses on some indicators including Apparent Oxygen Utilisation, Dissolved Oxygen, Nitrate, Oxygen Saturation, Phosphate, Sea Water Salinity, Sea Water Temperature, and Silicate; www.nodc.noaa.gov/OC5/WOA09/pr_woa09.html
- Marine Regions: The data source gives access to a standard list of marine georeferenced place names and areas including EEZ; www.marineregions.org

- myOceans: The data source gives access to a number of environmental variables. In particular, D4Science focuses on some indicators including ice concentration, ice thickness, ice velocity, mass concentration of chlorophyll in sea water, meridional velocity, mole concentration of dissolved oxygen in sea water, mole concentration of nitrate in sea water, mole concentration of phosphate in sea water, mole concentration of phytoplankton expressed as carbon in sea water, net primary production of carbon, salinity, sea surface height, temperature, zonal velocity, wind speed, and wind stress. http://www.myocean.eu

3.2.2.3 STATISTICAL DATA

The following databases and information systems have been integrated to give access to statistical datasets:

- IRD Datasets: The UMR EME/Observatoire Thonier SDMX Registry and Repository exposes the Sardara database that contains tuna captures data from several countries, aggregated according to CWP statistical squares (1'x1' or 5'x5') and the ObServe database that contains tuna and bycatches captures observed by scientific observers on-board of French industrial purse seiners.
- Codelists: A set of SDMX Codelists either directly accessed from the FAO Registry, or manually uploaded through the facility developed in the context of ICIS.

3.2.2.4 OTHER DATA

The following databases and information systems have been integrated to give access to other resources including papers:

- Aquatic Commons: The data source offers access to thematic material covering natural marine, estuarine/brackish and fresh water environments; aquaticcommons.org
- BHL: The Biodiversity Heritage Library data source offers access to legacy literature of biodiversity held by a consortium of natural history and botanical libraries; www.biodiversitylibrary.org
- Bioline: The Bioline International data source offers access to open access quality research journals published in developing countries; www.bioline.org.br
- CEEMar: The Central and Eastern European Marine Repository data source offers material covering marine, brackish and fresh water environments; www.ceemar.org/dspace
- DataCite: The data source offers access to the same service whose mission is to give access to research data; www.datacite.org
- DBPedia: The knowledge base results from Wikipedia. It contains over 4 millions things including persons, places, creative works, organisations, species and diseases; dbpedia.org/About
- DRS: The data source at National Institute of Oceanography offers institutional publications including journal articles and technical reports; drs.nio.org/drs
- Dryad: The data source offers access to the same service whose mission is to give access to research data underlying research publications; datadryad.org
- FactForge: The knowledge base results from the integration of a number of datasets including DBPedia, WordNet, Geonames, and Freebase; factforge.net
- FAO Factsheets: The data source gives access to the Aquatic Species Fact Sheets developed by the same FAO programme; www.fao.org/fishery/fishfinder

- FAO FLOD: A semantic knowledge base hosted by FAO containing a dense network of relationships among the major entities of the fishery domain, including marine species, water areas, land areas, and exclusive economic zones; www.fao.org/figis/flod
- iMarine TLO Warehous: The warehouse integrates information from FishBase, WoRMS, ECOSCOPE, FLOD and DBPedia by using the same top-level ontology developed for the marine domain. It currently contains approximately 3 millions of triples about more than 40,000 entities including marine species, ecosystems, water areas, and vessels; http://www.ics.forth.gr/isl/MarineTLO
- Nature: The data source offers access to the articles published by nature.com.
- OceanDocs: The data source offers research and publication materials in Marine Science by aggregating content form 256 repositories; www.oceandocs.net
- OpenAIRE: The data source give access to the publications aggregated by the same European funded project; www.openaire.eu
- PANGAEA: The data source offers georeferenced data from earth system research via OAI-PMH. The system guarantees long-term availability of its content through a commitment of the operating institutions. The aggregated repositories are 475; www.pangaea.de
- PenSoft Journals: The data source gives access to a number of open-access journals. In particular, iMarine focuses on BioRisk, Comparative Cytogenetics, International Journal of Myriapodology, Journal of Hymenoptera Research, MycoKeys, Nature Conservation, NeoBiota, PhytoKeys, Subterranean Biology, and ZooKeys.
- SmartFish: The SmartFish Chimaera knowledge base offers a unified and integrated view on three marine fisheries information sources, i.e. FIRMS - an international knowledge base including fisheries and resource from West Indian Ocean; StatBase - a statistical database containing statistics provided by West Indian Ocean countries; and WIOFish - a regional knowledge base on West Indian Ocean Fisheries.
- WHOAS: The data source offers the production of Woods Hole scientific community including articles and data sets; www.mblwhoilibrary.org/services/whoas-repository-services
- YAGO2: The knowledge base extends the YAGO knowledge base by anchoring entities, facts and events in time and space. The knowledge base is built from Wikipedia, GeoNames and WordNet and contains more than 440 million facts about 9.8 million entities.

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.

The part of the production resources allocated to QA task is summarized in Table 9. 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.

Table 9. Quality Assurance Infrastructure resources

Site	VMs
CNR	14
NKUA	6
Total	**20**

4 MONITORING AND ACCOUNTING

As reported in [10], in iMarine it is exploited a comprehensive suite of monitoring and accounting tools including gCube tecnologies and third party facilities like Nagios [19], Ganglia, and Google Analytics.

Part of the third party tools was hosted and operated by CERN on their premises up to May 2014, namely Nagios. From June 2014 on, the management and operation of such tools was transferred to CNR.

4.1 INFRASTRUCTURE USAGE INDICATORS

A number of indicators are reported below to describe the actual use of infrastructrue and its services performed during the reporting period.

4.1.1 PORTAL USERS

Figure 7 reports on the portal sessions performed during the reporting period. A session is the period time a user is actively engaged with the facilities offered through the project portal. In the reporting period a total of 14,419 sessions occurred with an average of approximately 900 sessions per month.

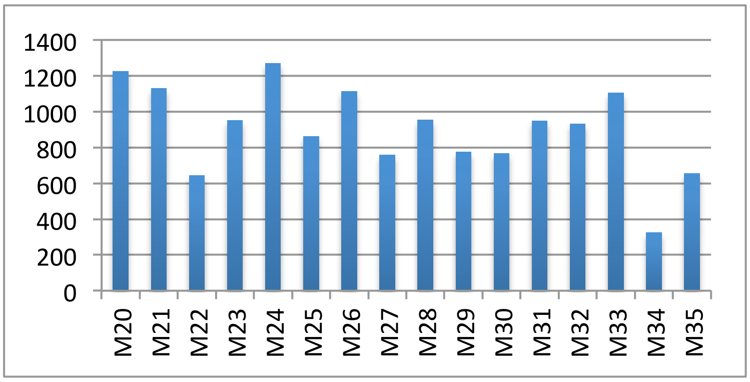

Figure 7. Portal sessions by month

Figure 8 reports on the spatial distribution of the sessions. From this picture it emerges that the largest part of sessions originates from countries belonging to the Souther Europe region (68% of the sessions), however the overall coverage is almost global.

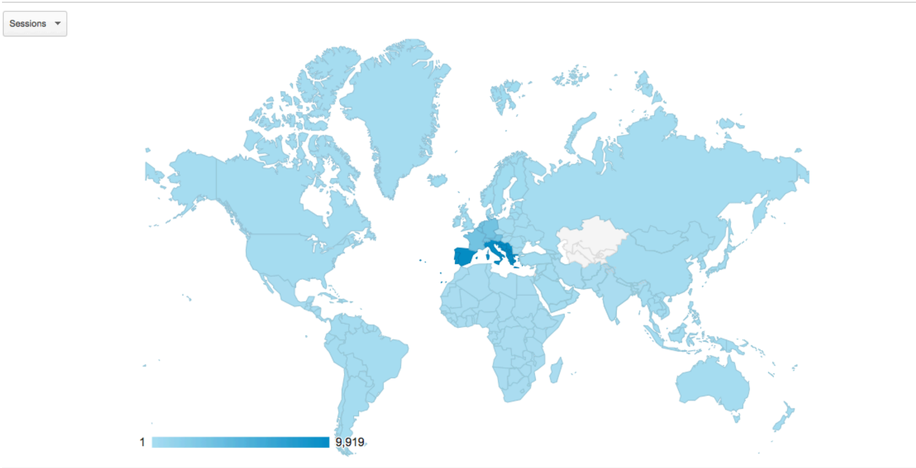

Figure 8. Portal sessions by sub continent region

Figure 9 reports on the number of users that have had at least one session during the reporting period. A total of 2,639 users have been served through the portal with an average of 221 users per month.

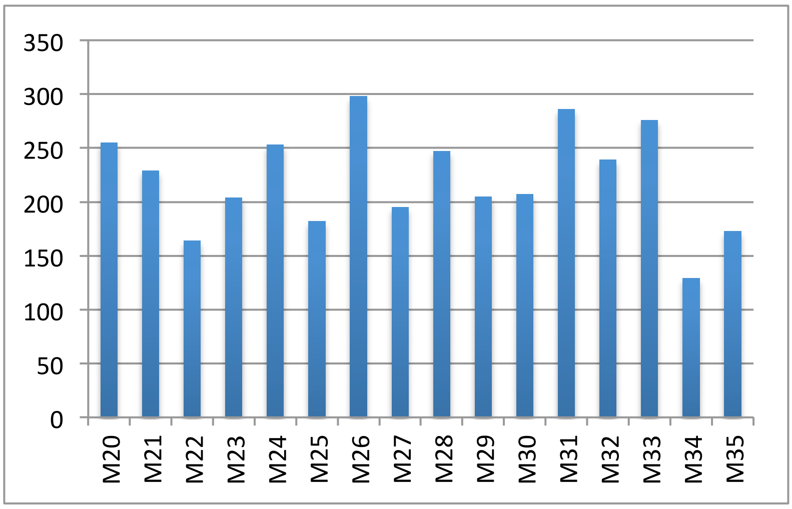

Figure 9. Portal users by month

4.1.2 HOSTING NODES EXPLOITATION

As presented in Section 3.1.2 the number of gHNs running in the infrastructure at the end of the reporting period was in the order of one hundred. These gHNs are exploited for the deployment of other gCube services to serve particular VOs and VREs. 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 System. Given that, it is important to report the distribution of Running Instances as they are registered on each VOs' Information System (see Figure 10).

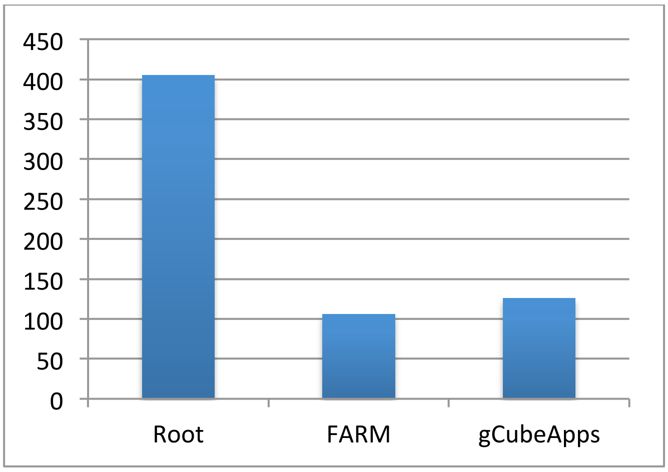

Figure 10. Running Instances by VO

These figures are almost equivalent to that presented in the previous period. 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).

4.2 INFRASTRUCTURE AVAILABILITY

As previously reported, the nodes and services of the infrastructure were 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.

In the case of gCube Hosting Nodes we have an average of 99.5% of services up (Figure 11 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
dl25.di.uoa.gr	97.966% (97.966%)	2.034% (2.034%)	0.000% (0.000%)	0.000%
dl27.di.uoa.gr	97.980% (97.980%)	2.020% (2.020%)	0.000% (0.000%)	0.000%
dl28.di.uoa.gr	97.980% (97.980%)	2.020% (2.020%)	0.000% (0.000%)	0.000%
mlhec4.di.uoa.gr	97.964% (97.964%)	2.036% (2.036%)	0.000% (0.000%)	0.000%
mln11.di.uoa.gr	97.975% (97.975%)	2.025% (2.025%)	0.000% (0.000%)	0.000%
node1.p.d4science.research-infrastructures.eu	99.955% (99.955%)	0.045% (0.045%)	0.000% (0.000%)	0.000%
node10.p.d4science.research-infrastructures.eu	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
node11.d4science.org	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
node11.p.d4science.research-infrastructures.eu	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
node12.d4science.org	99.982% (99.982%)	0.018% (0.018%)	0.000% (0.000%)	0.000%
node12.p.d4science.research-infrastructures.eu	99.964% (99.964%)	0.036% (0.036%)	0.000% (0.000%)	0.000%
node13.d4science.org	99.982% (99.982%)	0.018% (0.018%)	0.000% (0.000%)	0.000%
node13.p.d4science.research-infrastructures.eu	99.964% (99.964%)	0.036% (0.036%)	0.000% (0.000%)	0.000%
node14.d4science.org	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
node15.d4science.org	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
node16.d4science.org	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
node16.p.d4science.research-infrastructures.eu	99.978% (99.978%)	0.022% (0.022%)	0.000% (0.000%)	0.000%
node17.p.d4science.research-infrastructures.eu	99.977% (99.977%)	0.023% (0.023%)	0.000% (0.000%)	0.000%
statistical-manager1.d4science.org	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
statistical-manager2.d4science.org	99.997% (99.997%)	0.003% (0.003%)	0.000% (0.000%)	0.000%
statistical-manager3.d4science.org	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
statistical-manager4.d4science.org	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
Average	99.504% (99.504%)	0.496% (0.496%)	0.000% (0.000%)	0.000%

Figure 11. Availability Report for GHNs

The availability of the services contributing to form part of the iMarine SDI, namely GeoNetwork and GeoServer, is reported in Figure 12. In this case, the average of services up is 99.99%.

Hostgroup 'Geoserver' Host State Breakdowns:

Host	% Time Up	% Time Down	% Time Unreachable	% Time Undetermined
geonetwork.d4science.org	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
geoserver.d4science.org	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
geoserver2.d4science.org	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
geoserver3.d4science.org	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
geoserver4.d4science.org	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
Average	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%

Figure 12. Availability Report for iMarine SDI (GeoNetwork and GeoServer instances)

The availability of the MongoDB cluster, the third party technology exploited to realise the gCube File-based storage, is reported in Figure 13. In this case, the average of services up is 99.99%.

Hostgroup 'Mongo Servers' Host State Breakdowns:

Host	% Time Up	% Time Down	% Time Unreachable	% Time Undetermined
node58.p.d4science.research-infrastructures.eu	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
node67.p.d4science.research-infrastructures.eu	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
node73.p.d4science.research-infrastructures.eu	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
node80.p.d4science.research-infrastructures.eu	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
Average	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%

Figure 13. Availability Report for MongoDB Cluster

The availability of the Cassandra cluster, the third party technology supporting the operation of the gCube social networking facilities, is reported in Figure 14. In this case, the average of services up is 99.99%.

Hostgroup 'Cassandra' Host State Breakdowns:

Host	% Time Up	% Time Down	% Time Unreachable	% Time Undetermined
node1.p.cassandra.research-infrastructures.eu	99.997% (99.997%)	0.003% (0.003%)	0.000% (0.000%)	0.000%
node2.p.cassandra.research-infrastructures.eu	99.998% (99.998%)	0.002% (0.002%)	0.000% (0.000%)	0.000%
Average	99.997% (99.997%)	0.003% (0.003%)	0.000% (0.000%)	0.000%

Figure 14. Availability Report for Cassandra Cluster

The availability of the Hadoop cluster, the third party technology exploited by a series of gCube services to execute various tasks in a distributed way, is reported in Figure 15. In this case, the average of services up is 100%.

Hostgroup 'hadoop-cluster' Host State Breakdowns:

Host	% Time Up	% Time Down	% Time Unreachable	% Time Undetermined
jobtracker.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
logstash.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
node10.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
node11.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
node12.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
node13.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
node2.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
node3.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
node4.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
node5.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
node6.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
node7.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
node8.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
node9.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
quorum1.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
quorum2.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
quorum3.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
quorum4.t.hadoop.research-infrastructures.eu	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
Average	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%

Figure 15. Availability Report for the Hadoop Cluster

4.2.1 INFRASTRUCTURE DOWNTIMES

Figure 16 presents the total number of downtimes declared by the sites that host the iMarine resources of the D4Science infrastructure.

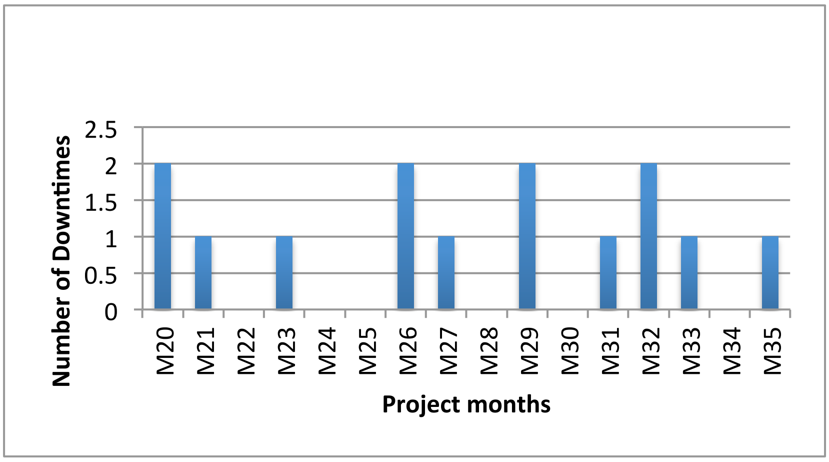

Figure 16. Infrastructure resources downtimes by month

It should be mentioned that these "downtimes" were caused either by scheduled network interventions or by infrastructure upgrades. 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. In average, in the reporting period there has been 0.875 downtimes per month. The downtimes lasted in average 6.5 hours (infrastructure upgrades took also 8 hours and there was 8 of such upgrades in the reporting period), so the production infrastructure was never affected by long unavailability periods and each single VRE was unavailable for just a short period of time.

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 [18]. 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 [9] together with detailed information about the procedures, the people involved, and the workflow of the different steps.

A total number of ***237 tickets*** were submitted and closed during the reporting period. The 53% (126 tickets) of the managed ticket was high priority incidents. Figure 17 gives detailed information about the total number of submitted tickets by submission month (high priority tickets are highlighted).

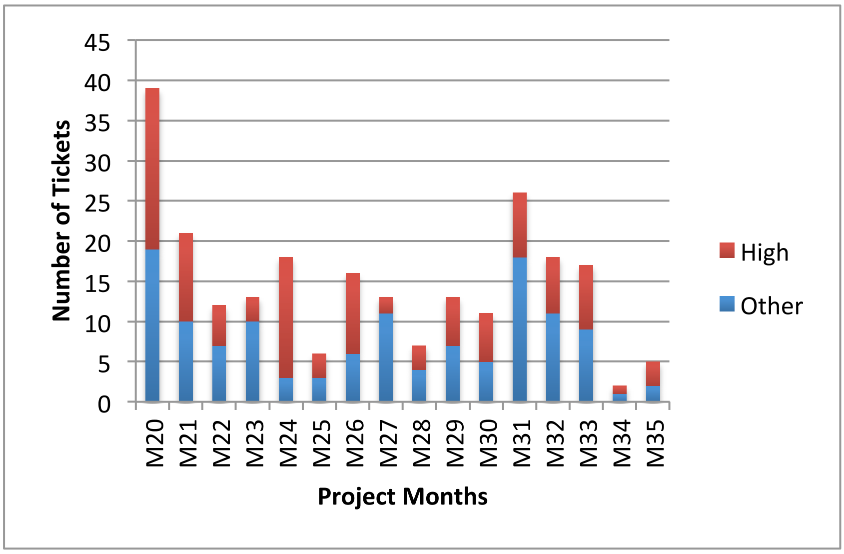

Figure 17. Incident Tickets by Priority

Figure 18 shows the distribution of the recorded incident tickets across the support scopes. From this figure it emerges that the large majority of the tickets were not specific to a particular VRE, rather were common to all existing VREs. This is mainly due to the fact that the VREs shares many services. The 70% (166 tickets) of tickets were not specific to any VRE, while the 10% (25 tickets) of tickets was reported by referring to FARM VREs and the 19% (46) of tickets was reported by referring to gCubeApps VREs.

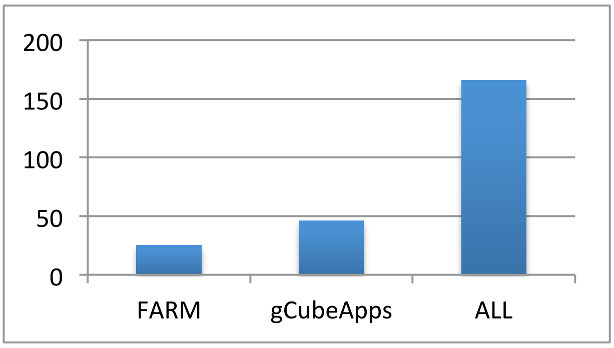

Figure 18. Incident tickets by VO

6 CONCLUDING REMARKS

This deliverable described the activities resulting from the operation of the iMarine infrastructure in the period June 2013 – September 2014.

In the reporting period the infrastructure was called to support the deployment and operation of two Virtual Organisations and twenty Virtual Research Environments. Overall, these VOs and VREs have served more than 750 users. In average, it has been observed an increase of +24% (223 at Jun '13 vs 278 at Sep '14) in the number of users served by FARM VREs and of +129% (214 at Jun '13 vs 492 at Sep '14) in the number of users served by gCubeApps VREs. In additiont to that, more than 400 additional users have been served via other portals (e.g. EGIP) and VREs (e.g. ENVRI) developed and operated by relying on gCube and iMarine technologies and procedures.

During the reporting period the infrastructure operation was based on a total of 11 gCube releases. The deployment of these releases lead to some downtime properly announced. In average, there has been 0.875 downtimes per month. The downtimes lasted in average 6.5 hours (infrastructure upgrades took also 8 hours and there was 8 of such upgrades in the reporting period), so the production infrastructure was never affected by long unavailability periods and each single VRE was unavailable for just a short period of time.

Through the iMarine portal a total of 14,419 sessions were served with an average of approximately 900 sessions per month. These sessions were started by a total of 2,639 users with an average of 221 users per month.

A total number of 237 tickets were submitted and closed during the reporting period. The 53% (126 tickets) of the managed tickets was high priority incidents. The 70% (166 tickets) of tickets were not specific to any VRE, while the 10% (25 tickets) of tickets was reported by referring to FARM VREs and the 19% (46) of tickets was reported by referring to gCubeApps VREs. This is quite normal since there are services that are not specific to any VRE as well as there are services that are shared across more than one VRE thus a malfunction can not attributed to a specific VRE.

REFERENCES

[1] L. Candela. Virtual Research Environments Plan. D6.1 iMarine Project Deliverable. January 2012
[2] L. Candela, A. Ellenbroek, P. Pagano. Virtual Research Environments Activity Report. iMarine D6.5 Project Deliverable. September 2014
[3] D4Science Consortium. The D4Science.org Web site http://www.d4science.org/
[4] EuBrazilOpenBio Consortium. The EuBrazilOpenBio Web Site http://www.eubrazilopenbio.eu
[5] ENVRI Consortium. The ENVRI Web site http://envri.eu/
[6] iMarine Consortium. The iMarine Web site http://www.i-marine.eu/
[7] iMarine Consortium. The Ecosystem Approach Community of Practice Wiki page http://wiki.i-marine.eu/index.php/Ecosystem_Approach_Community_of_Practice_Overview
[8] iMarine Consortium. Ecosystem Approach Community of Practice: Requirements http://wiki.i-marine.eu/index.php/Ecosystem_Approach_Community_of_Practice:_Requirements
[9] A. Manzi. iMarine Data e-Infrastructure Plan. iMarine D5.1 Project Deliverable. January 2012
[10] A. Manzi, L. Candela. iMarine Data e-Infrastructure Operation Report. iMarine D5.3 Project Deliverable. June 2013
[11] The UMD Distribution Web Site http://repository.egi.eu/category/umd_releases/
[12] The EMI Project Web Site http://www.eu-emi.eu/
[13] The VOMS server admin page https://voms.research-infrastructures.eu:8443/voms/d4science.research-infrastructures.eu
[14] The EGI Infrastructure Web site http://www.egi.eu/
[15] The VENUS-C project Web Site http://www.venus-c.eu
[16] The MS Azure Web Site http://www.windowsazure.com/
[17] The MyEGI Web Site https://tomtools.cern.ch/confluence/display/SAM/MyEGI
[18] The iMarine incident management procedure http://wiki.i-marine.eu/index.php/Procedure_Infrastructure_Incident_Management
[19] The Nagios Web Site http://www.nagios.org/