Project acronym	D4Science-II
Project full title	Data infrastructure ecosystem for science
Project No	239019

Knowledge Ecosystem Operation Report

Deliverable No DSA1.2a

September 2010

SEVENTH FRAMEWORK PROGRAMME
Research Infrastructures

INFRA-2008-1.2.2: Scientific Data Infrastructures

DOCUMENT INFORMATION

Project	
Project acronym:	D4Science-II
Project full title:	Data Infrastructures Ecosystem for Science
Project start:	1st October 2009
Project duration:	24 months
Call:	INFRA-2008-1.2.2: Scientific Data Infrastructures
Grant agreement no.:	239019
Document	
Deliverable number:	DSA1.2a
Deliverable title:	Knowledge Ecosystem Operation Report
Contractual Date of Delivery:	September 2010
Actual Date of Delivery:	15 October 2010
Editor(s):	P. Andrade
Author(s):	P. Andrade, L. Candela, A. Manzi, P. Pagano
Reviewer(s):	A. Ellenbroek
Participant(s):	CERN, CNR
Work package no.:	SA1
Work package title:	Knowledge Ecosystem Operation
Work package leader:	CERN
Work package participants:	CERN, CNR, ENG, FAO, FIN, NKUA
Est. Person-months:	3
Distribution:	Public
Nature:	Report
Version/Revision:	1.0
Draft/Final	Final
Total number of pages: (including cover)	25
Keywords:	Ecosystem, Infrastructure, VREs, Status

CHANGE LOG

Reason for change	Issue	Revision	Date
Initial table of contents	0	1	16 Aug
Version for internal review	0	3	20 Sep
Version for official review	1	0	5 Oct
Reviewed version	1	0	6 Oct

DISCLAIMER

This document contains description of the D4Science and D4Science-II project findings, work and products. Certain parts of it might be under partner Intellectual Property Right (IPR) rules so, prior to using its content please contact the consortium head for approval.
E-mail: info@d4science-ii.research-infrastructures.eu

In case you believe that this document harms in any way IPR held by you as a person or as a representative of an entity, please do notify us immediately.

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 in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of D4Science-II consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 27 Member States of the 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/)

D4Science-II is a project partially funded by the European Union

LIST OF ABBREVIATIONS

D4Science-II	Data infrastructure ecosystem for science
DoW	Description of Work
EC	European Commission
EM	Environmental Monitoring
FARM	Fishery and Aquaculture Resource Management
gHN	gCube Hosting Node
IS	Information System
JRA	Joint Research Activity
MS	Messaging System
NA	Network Activity
ROC	Regional Operations Centres
SA	Service Activity
VO	Virtual Organisation
VRE	Virtual Research Environment
WSRF	Web Services Resource Framework

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This deliverable reports the activities carried out by the D4Science-II Service Activity (SA) to deploy the project production ecosystem. This ecosystem can be defined as the set of hardware, software, data collections, and procedures that together provide reliable collaboration environments to the D4Science-II user communities.

This document documents the work carried out during the first year of the project to deploy and maintain the production ecosystem. This includes: (1) the description of the infrastructure, sites, and nodes allocated to the ecosystem, (2) the results of applying the procedures defined for deployment, certification, monitoring, accounting, support, and finally (3) the description of the environments deployed to satisfy the requirements of the project user communities.

EXECUTIVE SUMMARY

The objective of the D4Science-II Service Activity is to make available and maintain a distributed ecosystem to support the activities of the project's user communities. These communities access the ecosystem via the exploitation of Virtual Research Environments (VREs) that satisfy concrete needs of the communities' Virtual Organizations (VOs).

This report documents the work done to establish the D4Science ecosystem during the first year of the D4Science-II project. It describes the different infrastructures that compose the ecosystem, the activities carried out to manage the infrastructure nodes, the activities carried out for monitoring and accounting of the infrastructure, the activities carried out to provide efficient support to the ecosystem users, and finally the VOs and VREs deploy to satisfy the D4Science-II scientific scenarios.

The D4Science ecosystem is composed by many infrastructures being the most important, and mandatory for the overall orchestration of the ecosystem, the D4Science infrastructure. Different from the other infrastructures whose resources are managed by external organizations, the D4Science infrastructure resources are managed by the project partners. As a consequence this report describes the ecosystem in general but provides detailed information only for the D4Science infrastructure itself.

To describe the work carried out to deploy the D4Science ecosystem it is important to introduce the scientific scenarios served by the ecosystem and the infrastructures that compose the ecosystem. The ecosystem scenarios are: INSPIRE, DRIVER, FCPPS, AquaMaps, and ICIS while the ecosystem infrastructures are: D4Science, EGEE, AquaMaps, GENESI-DR, DRIVER, and INSPIRE.

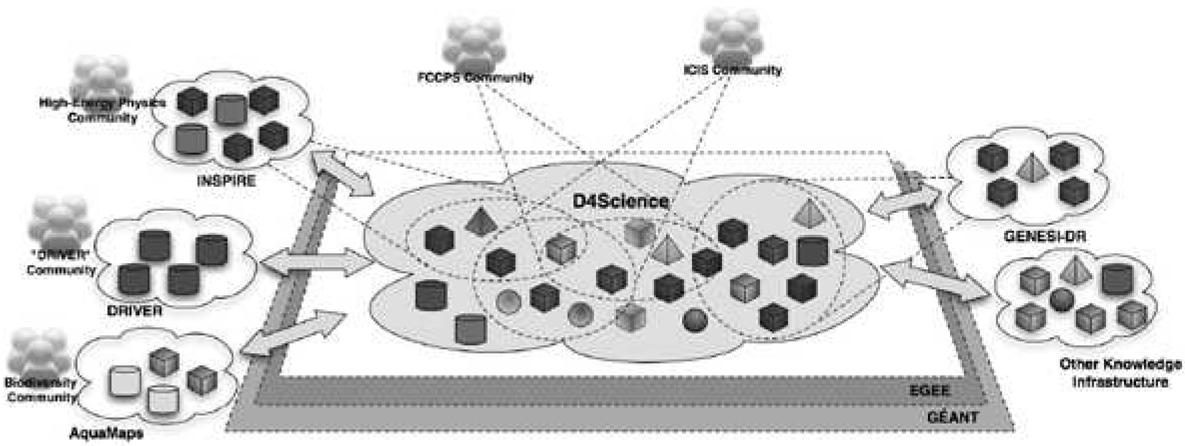

Figure 1 - D4Science ecosystem scenarios and infrastructures

From the Service Activity point of view, and as defined in the project Description of Work, the D4Science ecosystem must serve the different scientific scenarios according to the following milestones:

- Jan 2010 (M4) – AquaMaps, FCPPS, and ICIS;
- Sep 2010 (M12) – AquaMaps, FCPPS, ICIS, INSPIRE, and DRIVER;
- Mar 2011 (M18) – AquaMaps, FCPPS, ICIS, INSPIRE, and DRIVER;
- Sep 2011 (M24) – AquaMaps, FCPPS, ICIS, INSPIRE, and DRIVER.

Since these scientific scenarios have direct dependencies on certain infrastructures, the integration of the different infrastructures in the ecosystem is also affected.

During the first year the focus was on setting up VOs and VREs for the AquaMaps, FCPPS, and ICIS **scenarios**. These scenarios required the integration of the D4Science, EGEE, and AquaMaps **infrastructures** in the ecosystem. This work was accomplished, as planned, on January 2010. Since then, these environments have been maintained and new developments started on the DRIVER and INSPIRE scenarios which are now about to be officially deployed in the production ecosystem.

As originally defined in deliverable DSA1.1a the D4Science infrastructure gathers a number of **nodes** provided by SA1 work package members. In total four sites contributed to the infrastructure: CNR, FAO, FIN, and NKUA. At the end of the first year these sites provide a total of 39 physical machines, offering 503 GB RAM, 35,4 TB disk space, and 148 processor cores. These nodes were exploited to run the gCube system delivered by the D4Science SA3 work package and the gLite middleware as released by the EGEE project. As a result the infrastructure makes now available 74 gCube nodes and has access to 4881 gLite nodes.

The management of the infrastructure was facilitated by the definition and implementation of clear procedures for **monitoring** and **accounting**. A number of monitoring tools were deployed allowing different infrastructures roles to visualize the status of their resources and to be actively notified when problems occurred. An accounting tool was also put in production providing relevant statistics about the users' exploitation of the infrastructure.

A clear **support** procedure was also defined and put in production during the course of the year. This procedure ensured an efficient response to all incidents affecting the ecosystem operation. A total number of 168 tickets were submitted, 55,9% being of high priority. All tickets were properly closed and documented.

From the perspective of the ecosystem user different **environments** have been deployed and made available satisfying the requirements expressed by the different project scientific scenarios. In total 4 VOs and 8 VREs were made available (some were already dismissed).

1 NODES OPERATION

This section describes the hardware resources that compose the D4Science infrastructure, the core infrastructure of the D4Science ecosystem, and the activities carried out during the project lifetime to operate such nodes. These activities include the deployment and upgrade of gCube and gLite in the infrastructure, the certification of the infrastructure nodes, and the management of infrastructure downtimes.

The D4Science infrastructure is organized in two node types:

- **gCube nodes** are the hardware resources able to run gCube services. The gCube software includes a particular web service container, the gCube Hosting Node (gHN), and a set of services and libraries that provide the functionality to create, manage, and exploit VREs;
- **gLite nodes** are computing and storage nodes running gLite software. gLite is the middleware provided by the EGEE project. By running gLite, these nodes provide core grid functionalities such as file-based storage, distributed computation of applications, etc. gLite nodes are exploited by gCube services which then provide higher level functionality through the D4Science VREs.

1.1 Hardware Resources

From the hardware point of view, there are currently four sites providing hardware resources to the production infrastructure:

- **CNR** – Pisa, Italy
- **FAO** – Rome, Italy
- **FIN** – Kiel, Germany
- **NKUA** – Athens, Greece

The following table provides detailed information about the contribution from each site.

	#	CPU	RAM GB	Disk TB	Cores
CNR	10	Sun v20z 2x AMD Opteron	80	1.5	20
	2	Supermicro Intel Core 2 Quad-core	8	8	8
	2	Sun X4140 2x AMD Opteron Six-core	128	1.2	24
	4	Sun X4100 2x AMD Opteron Quad-core	128	1.2	32
	4	Intel Core 2 Quad-core	28	15	16
	2	Sun X2100 AMD Opteron Dual-Core	16	1	4
	2	Asus: AMD Opteron 242	12	0.64	4
	2	Asus: AMD Athlon 64 X2 Dual-core	16	1	4
	1	Sun Cobalt LX50 Intel Dual Pentium III	3	0.15	2
FAO	1	2 Intel Xeon Quad-core	8	0.3	8
FIN	1	Intel Xeon E5504	12	1	4
NKUA	5	Dual Core Intel	40	2.9	10
	3	Intel Xeon	24	1.5	12
Total	**39**		**503**	**35.39**	**148**

Table 1 - Hardware resources by partner

The 39 physical machines provided so far are inline 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 SA2 and NA5 work packages.

1.2 Software Deployment

As introduced before, the resources allocated to the D4Science production infrastructure were exploited to host gCube and gLite nodes. Many of the hardware resources were virtualized in order to provide a higher number of gCube and gLite nodes. Due to the different nature of the two types of nodes the deployment and upgrade of software in these nodes is also distinct.

gCube Nodes

Of the four sites providing resources to the infrastructure, four sites deployed gCube nodes in their site. The common component to all gCube nodes is the gCube Hosting Node (gHN). The gHN is responsible to manage the deployed Running Instances (RIs) of the different gCube services.

In order to satisfy the requests from the project user communities for VOs and/or VREs, new nodes were continuously added to the infrastructure. **Error! Reference source not found.** shows this increase in the number of nodes (gHNs) and a constant number of RIs, reflecting the utilization from VOs and VREs of the available gHNs.

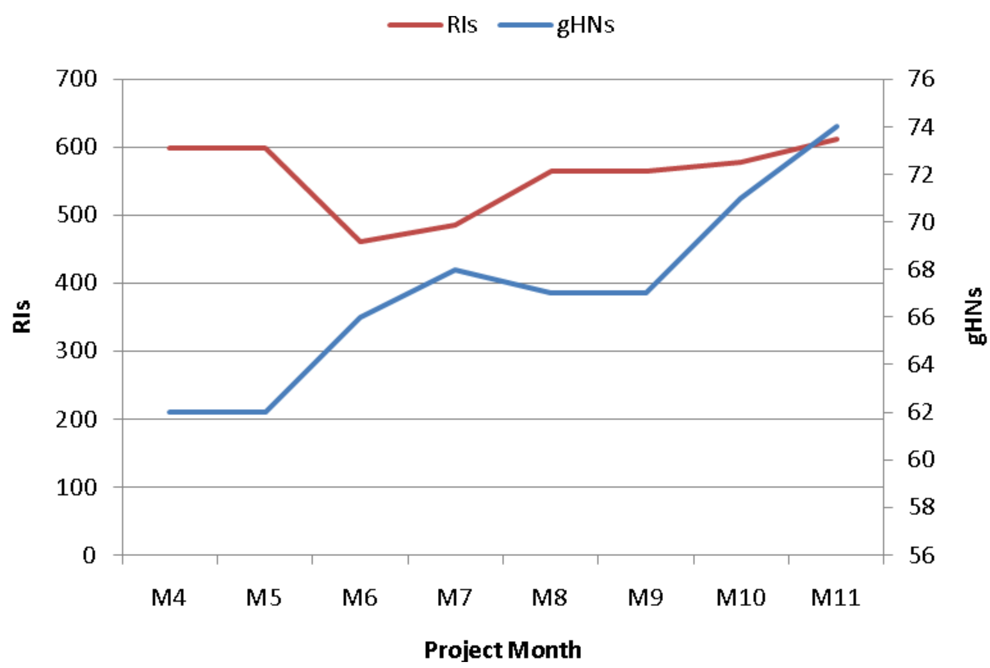

Figure 2 - Evolution of gHNs and RIs

The current gCube release deployed in production is release 2.0.0. During the first year of the project five major gCube releases were deployed in production (1.6.0, 1.7.0, 1.8.0, 1.9.0, 2.0.0). Due to the need to quickly fix critical defects affecting the production environment seven gCube maintenance releases were also deployed (1.6.1, 1.7.1, 1.7.2, 1.7.3, 1.7.4, 1.8.1, 1.9.1). Table 2 provides more information about the gCube major releases.

Release	Date	Integration Time	Subsystems	SAs
1.6.0	27 Jan	-	21	181
1.7.0	11 Feb	29	21	179
1.8.0	01 Jun	19	22	189
1.9.0	28 Jul	13	23	197
2.0.0	17 Sep	13	23	207

Table 2 - gCube releases deployed in the infrastructure

The chart below displays the information reported in Table 2 to highlight how the time needed to upgrade the infrastructure decreased even if the size of each release was always incresing from one release to the other. This behaviour reflects an automatization of the procedures established to integrate, test, and package new releases. The releases dates, spread by three different periods (early 2010, summer 2010, and September 2010), also reflect the evolution of the software with regular provision of code from the JRA work packages.

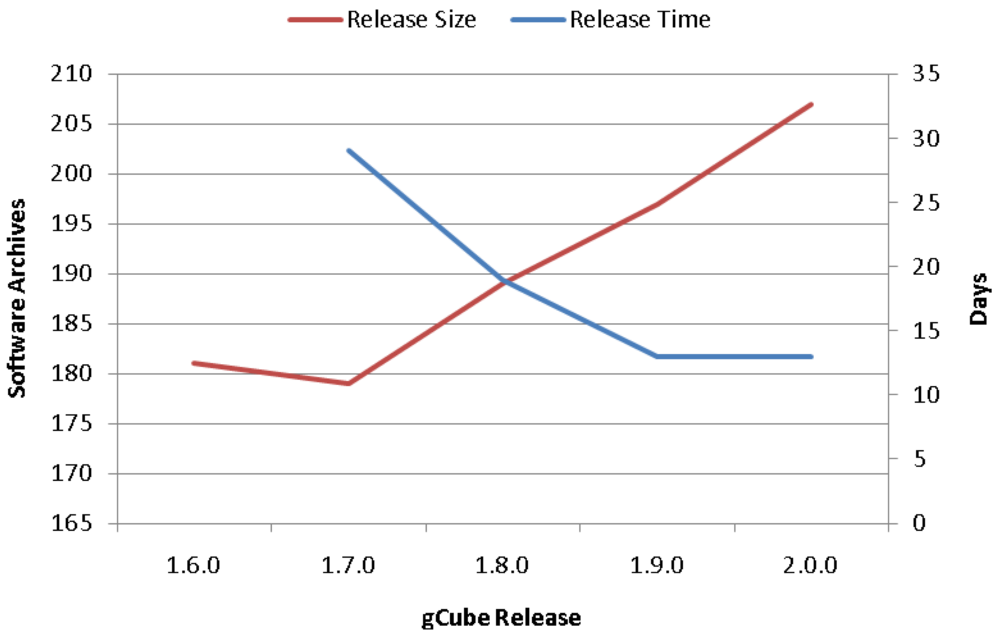

Figure 3 - gCube releases size and release integration time

This automatization of procedures is visible not only on the software integration tasks but also on the software deployment work. Figure 4 shows the number of days needed to deploy any new gCube release in the infrastructure. In average a normal release requires 4 days and maintenance release 1 day. It should be noted that usually the upgrade to a new release requires modifications to all infrastructure nodes. A full upgrade of hundreds of nodes in such short time is only possible due to the remote node management features provided by the gCube services.

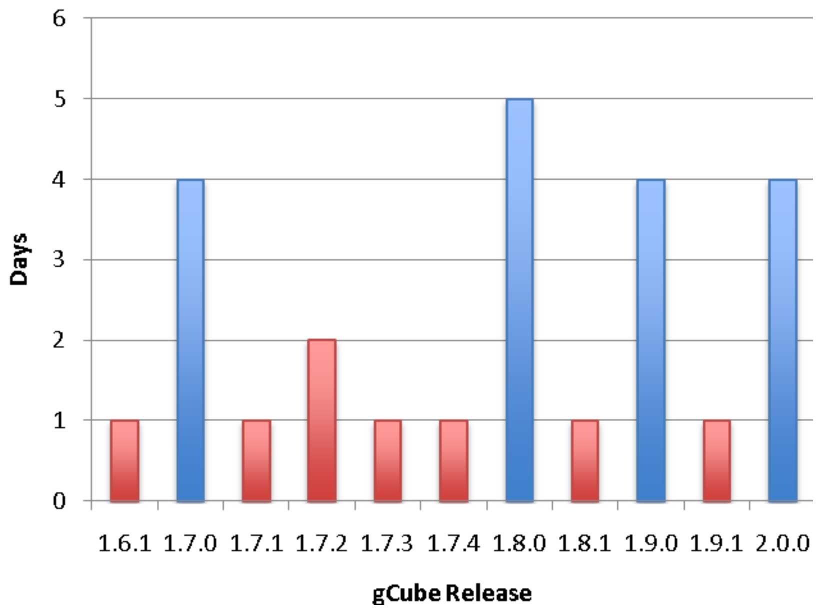

Figure 4 - gCube deployment time

gLite Nodes

Of the four sites providing resources to the infrastructure, two sites deployed gLite nodes in their site. Table 3 presents how the different gLite services have been distributed between these sites.

	CE	WN	sBDII	SE	WMS	LFC	VOMS	MyProxy
CNR	☑	☑	☑	☑	☑	☑	☑	☑
NKUA	☑	☑	☑	☑	☑	-	☑	-

Table 3 - gLite nodes per partner

The services marked in green represent a primary instance and in orange a secondary instance. All gLite sites provided by D4Science partners are also registered as sites of the EGEE project production infrastructure.

Besides the sites provided by project partners, other gLite sites of the EGEE production infrastructure also support the D4Science VO. These sites are: csTCDie, INFN Bari, INFN Trieste, and Taiwan.

In total these 7 sites provide access to 4881 gLite worker nodes distributed as follows:

- CNR – 10
- NKUA - 10
- csTCDie – 1089
- ESA – 10
- INFN Bari – 972
- INFN Trieste – 382
- Taiwan – 2408

Almost all gLite nodes provided by the above sites run gLite release 3.2. All patches for gLite 3.2 released during the last year were deployed in D4Science gLite nodes.

1.3 Nodes Certification

All nodes that compose the D4Science infrastructure must be certified before they can be exploited by the ecosystem.

gCube Nodes

Concerning the gCube nodes, there are no certification problems to report. All new nodes allocated to the infrastructure were kept up and running and registered in the gCube Information System with the correct activation status. With no certification problems to report all gHN containers were always available to host any service of the gCube system.

gLite Nodes

All nodes running gLite services were certified by the EGEE Regional Operation Centers. The two D4Science sites running gLite services (CNR and NKUA) kept their services up and running satisfying the service availability levels required by EGEE, therefore no new certification problems were raised.

1.4 Nodes Downtime

The following chart presents the total number of downtimes declared by the sites that compose the D4Sicence infrastructure.

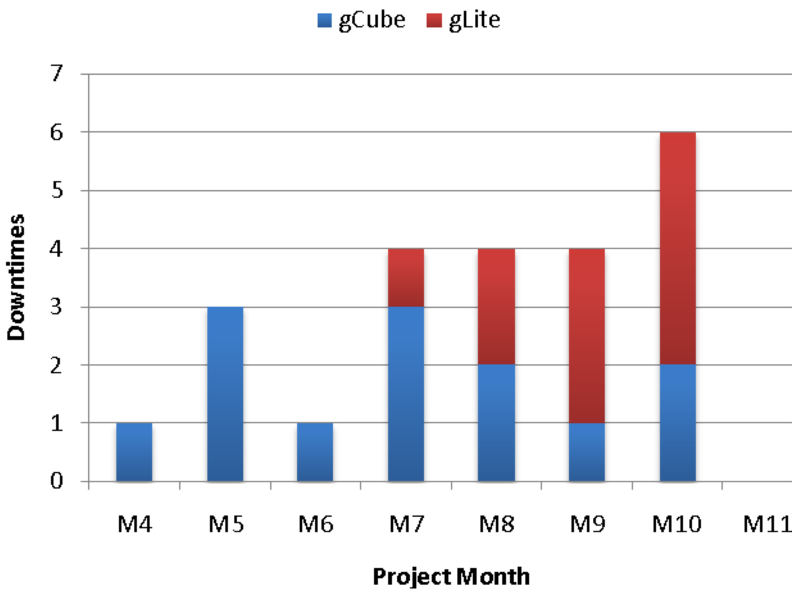

Figure 5 - Infrastructure nodes downtime

It should be mentioned that these downtimes were caused either by scheduled network interventions or by infrastructure upgrades. The downtimes never lasted for more than a few days so the production infrastructure was never affected by long unavailability periods. In downtimes motivated by infrastructure upgrades, even if full infrastructure downtimes were declared, only one VO was down at any time.

2 MONITORING AND ACCOUNTING

In a distributed infrastructure it is fundamental to have clear statistics about the infrastructure load and infrastructure usage. In D4Science a number of monitoring and accounting tools have been exploited. The monitoring and accounting of gLite nodes is based on tools provided by the EGEE project: GOCBD, SAM, and GSTAT.
The monitoring and accounting of gCube nodes is based on gCube services. These tools can be grouped in two categories: based on information collected by the gCube Information System (IS), and based on information produced/consumed by the gCube Messaging System (MS). While the gCube Messaging System provides tools for monitoring and accounting, the gCube Information system only provides tools for monitoring.
This section presents such tools and provides some statistics regarding the infrastructure load and infrastructure usage during the project first year.

2.1 Base Technology

As introduced before the monitoring and accounting tools for gCube nodes are based on two main gCube subsystems:

- gCube Information System
- gCube Messaging System

gCube Information System

The gCube IS subsystem allows the publication of descriptive information about VRE resources, the discovery of VRE resources based on descriptive information, and the real-time monitoring of VRE resources based on subscription/notification mechanisms. Heavily used by all functional layers of gCube, the IS is a replicated service whose instances communicate in peer- to-peer fashion to maximize availability, response time, and fault tolerance.

The IS plays a central role in a gCube-based Infrastructure by implementing the features supporting the publishing, discovery and real-time monitoring of the set of resources forming a gCube-based infrastructure. It acts as the registry of the infrastructure, i.e. all the resources are registered there and every service of the infrastructure must refer to it to dynamically discover the other infrastructure constituents.

The gCube IS subsystem is composed by different components:

- Components for production/publishing:
    - IS-Registry – This service supports the publishing/unpublishing of gCube resources. A gCube resource is advertised through its profile;
    - IS-gLiteBridge – This service supports the publishing/unpublishing of gLite resources gathered from the EGEE production infrastructure;
    - IS-Publisher – This library supports services in publishing/unpublishing groups of resource properties and registering/unregistering topics groups;
- Components for collection/storage:
    - IS-IC – This service aggregates the information published in the IS. Form a logical point of view it is a global registry but, because of the expected quality of service, it has been designed to support a federation model;
- Components for consumption/query:
    - IS-Client – This library supports services in retrieving information published in the IS. It supports the discovery of profiles and properties;
    - IS-Notifier – This service supports subscribing/unsubscribing of gCube services to topics produced by the other services;

  - IS-Monitoring – This component provides portlet interfaces to visualize the information gathered in the IS.

gCube Messaging System

The overall idea with a Messaging System is that it acts as a message mediator between message senders and message receivers. This mediation provides a new level of loose coupling for messaging. At a high level, messages are a unit of business information that is sent from one application to another via the MS. Applications send and receive messages via a MS using what are known as destinations. Messages are delivered to receivers that connect or subscribe to the messages.

This is the mechanism that allows for loose coupling between senders and receivers, as there is no requirement for each to be connected to the MS for sending and receiving messages. Senders know nothing about receivers and receives know nothing about senders – asynchronous messaging.

The gCube Messaging subsystem is composed by different components:

- Central Broker:
  - Message broker – receives and dispatches messages;
- Producer Side:
  - Messages – defines the messages to exchange;
  - Local producer – provides facilities to send messages from each node;
  - Node monitoring probes – produces monitoring info for each node;
  - Node accounting probes – produces accounting info for each node;
  - Portal accounting probes – produces accounting info for the portal;
- Consumer Side:
  - Messaging consumer – subscribes for messages from the message broker, checks metrics, stores messages, and notifies administrators;
  - Portal accounting portlet - displays portal accounting information;
  - Node accounting portlet - displays node accounting information.

2.2 Infrastructure Load

As presented in section 1.2 the number of gHNs running in the gCube nodes of the infrastructure was, during the project first year, in the order of 450-600. These gHNs are exploited for the deployment of other gCube services to serve particular VOs and VREs. Figure 6 shows the load on the infrastrucuture from these VOs and VREs. The chart clearly shows that, from the beginning of the project to the end of the first year, the average number of calls per hour per node almost increased by three.

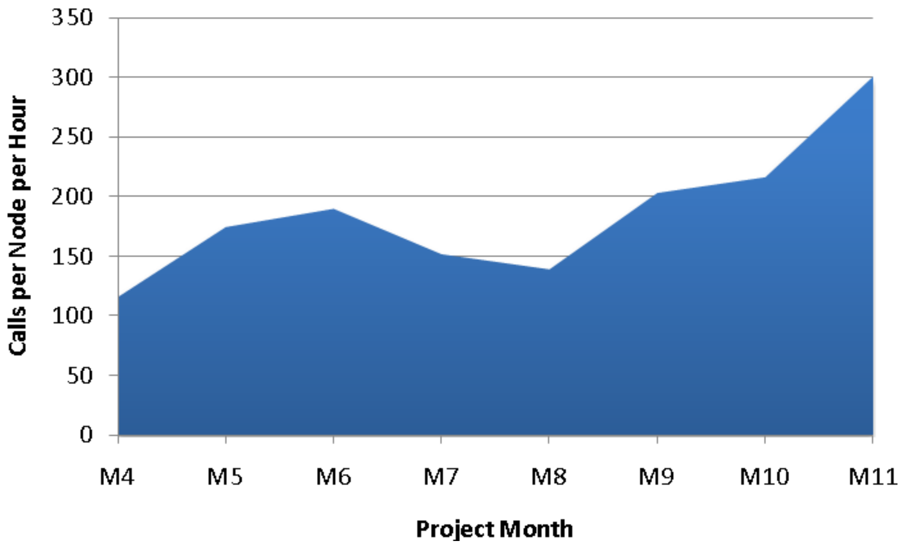

Figure 6 - Infrastructure load: average number of RIs calls

Due to the characteristics of the deployed monitoring tools it is possible to extract detailed information about the infrastructure load. If Figure 6 provide an overview of the average node utilization per hour, Table 4 depicts the total number of calls organized by gCube subsystem.

Subsystem	Calls
Information System	33454316
Content Management	251839
Metadata Management	17480
Search	17413
Application	15909
VRE Management	14965
Personalisation	2955
Messaging	2905
Execution	363
DIR	299
Annotation	272
Data Transformation	94

Table 4 - Infrastructure load: total number of RIs calls

From the table above it is clear the large impact of the gCube Information System components in the overall infrastructure load.

2.3 Infrastructure Usage

From the infrastructure exploitation view point, by relying on the deployed accounting tools, it is possible to verify a continuous usage of the infrastructure by its users.

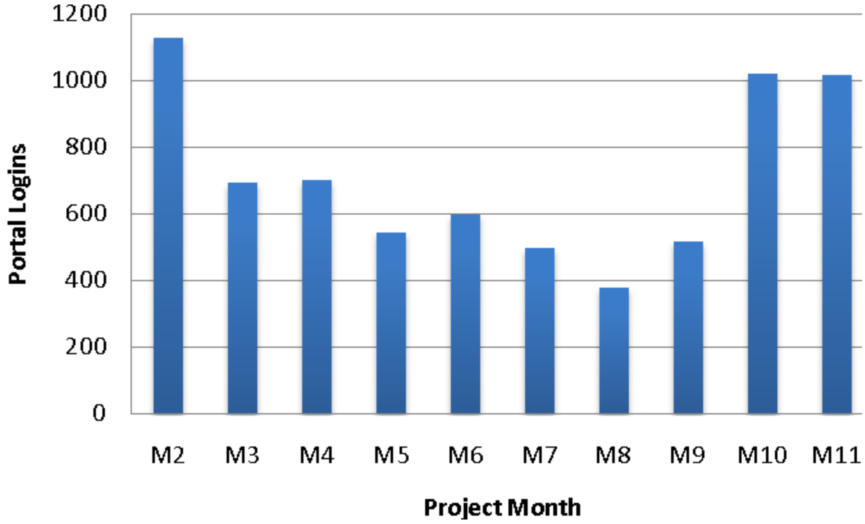

Figure 7 - Infrastructure usage: total number of logins

Figure 7 shows such usage by presenting the number of logins from the D4Science portal by month. As for the infrastructure load, more detailed statistics can also be produced for the infrastructure usage. Figure 8 for example explains how much the portal users have exploited the different VOs and VREs offered by the infrastructure. This percentage however only takes into consideration the actions of the users on a subset of the functionality offered by the portal: Search, Time Series, Archive Import, and Workspace.

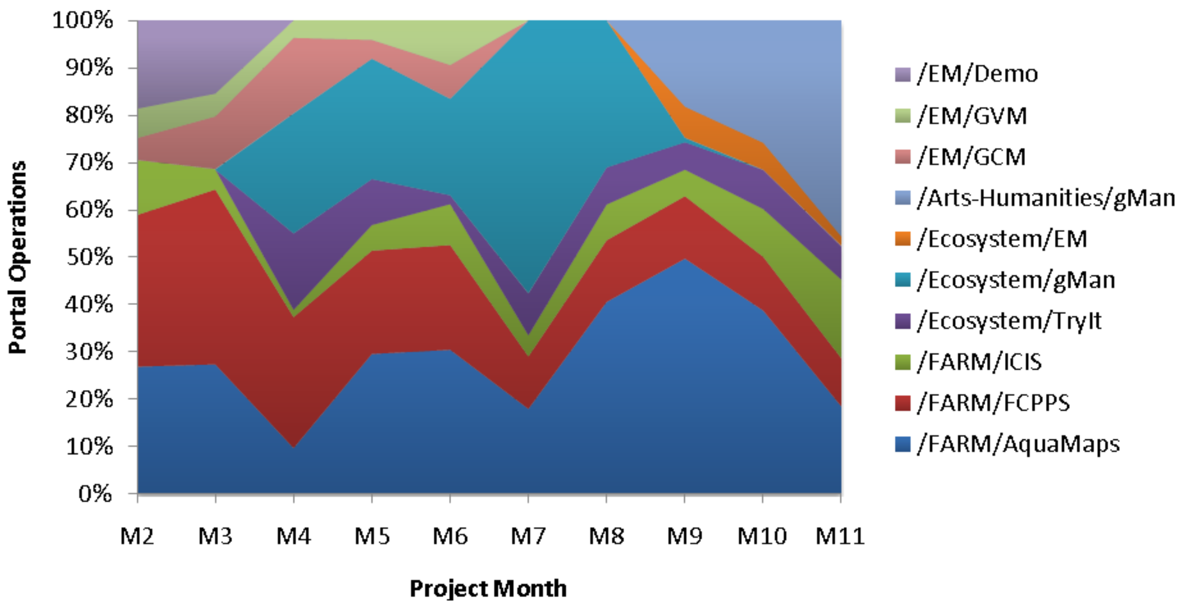

Figure 8 - Infrastructure usage: percentage of portal operation

3 PRODUCTION SUPPORT

This section describes the activities carried out to provide support to the operation and exploitation of the D4Science ecosystem by its different users: VO Administrators, VRE Managers, Site Managers, etc. The ecosystem support activity is based on the incident management procedure. This procedure follows the ITIL methodology for incident management and has been adopted since the beginning of the project.

The incident management procedure consists on the definition of different functional and hierarchical levels of support, making sure that all incidents are correctly solved according to their priority. The execution of the procedure is a task of the Support Team. The Support Team is composed by people from different activities of the project (SA1, SA2, JRA1, JRA2, and NA5) to make sure that enough knowledge is gathered to solve any type of incident. The main responsibilities of the Support Team are:

- Categorize, prioritize, and assign tickets;
- First line support (analyze assigned ticket and provide solution);
- Escalate tickets to TCom for 2nd level support (functional escalation);
- Escalate tickets to TB/PMB for a management decision (hierarchical escalation);
- Close and document tickets.

Detailed information about the procedures, the people involved, and the workflow of the different activities, is available in the D4Science infrastructure website [6] under the incident management section.

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

- gCube Nodes: All incidents related to the exploitation and deployment of gCube nodes are followed according to the incident management procedure;
- gLite Nodes: The incidents related to the usage of gLite nodes are followed according to the incident management procedure. The only exception concerns the incidents related to the deployment of gLite nodes, which are submitted directly to the EGEE Regional Operational Centre (ROC).

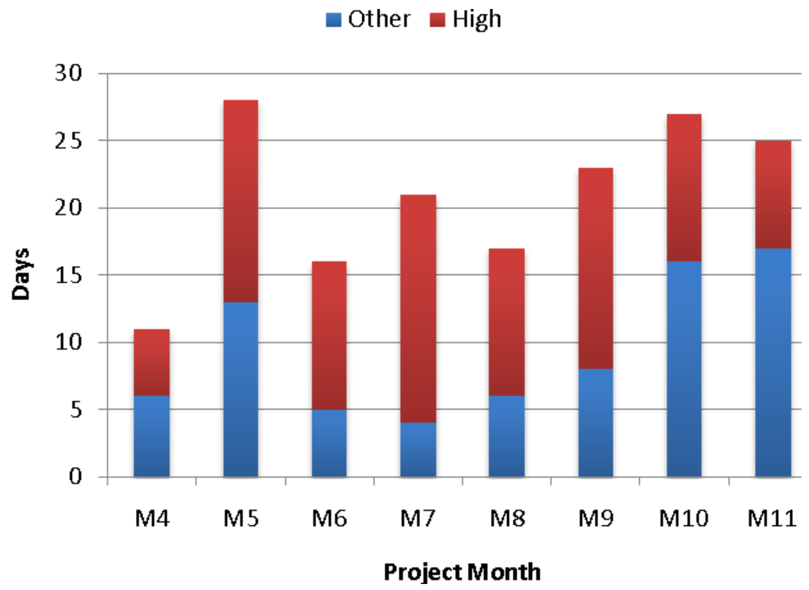

Figure 9 - Incident tickets by priority

A total number of 168 tickets were submitted during the project first year. From this total, 55,9% (94 tickets) were high priority incidents. Figure 9 and Figure 10 provide detailed information about the total number of submitted tickets, the number of high priority tickets, and the average number of days the tickets remained opened.

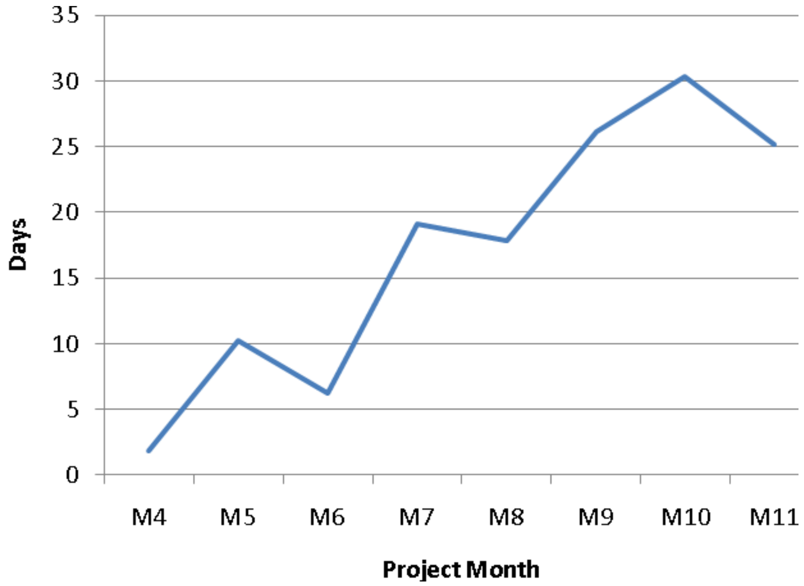

Figure 10 - Incident tickets average resolution time

The undesired increase in the number of days required to close incident tickets can be justified be the smaller number of maintenance releases prepared during the second part of the year. As presented in Table 2 (section 1.2) for releases 1.6.0 and 1.7.0 (from project M1 to M6) five maintenance releases were created while for releases 1.8.0, 1.9.0, and 2.0.0 (from project M7 to M12) only two maintenance releases were created.

Concerning the affected environments, Figure 11 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). Such higher number of incidents is expected as these VREs are the most exploited of the infrastructure.

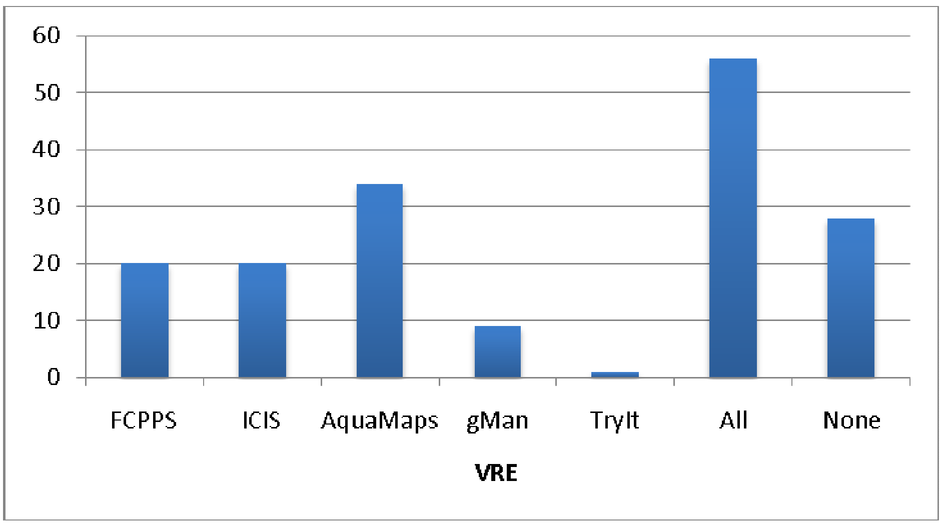

Figure 11 - Incident tickets by VRE

4 ECOSYSTEM 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 scenario.

The set of VOs and VREs the D4Science-II ecosystem 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 five scientific application scenarios the project is focused on – i.e. INSPIRE, DRIVER, AquaMaps, FCPPS and ICIS. These requirements are documented in "Communities Practices and Requirements" [1]. Specified requirements are carefully analysed and transformed into a deployment plan as documented in "VOs and VREs planning" [2]. This development plan is then implemented by a dedicated team and a living report is produced to document the activities performed [3]. In addition to the five scenarios introduced above, other VOs and VREs have been deployed 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 from January 2010 to September 2010[1]. During the period from January 2010 to September 2010 four Virtual Organisations have been deployed and operated in the context of the D4Science-II infrastructure:

- The **FARM** Virtual Organisation has been 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, and the dynamic generation of biodiversity maps and species probability maps.
- The **Ecosystem** Virtual Organisation has been created to serve the needs of any community interested in exploiting the D4Science-II Ecosystem capabilities. It is not conceived to serve the needs of a specific community focusing on a research topic. It aims at promoting a cross-fertilisation among the various communities partaking to it. The pool of resources is more heterogeneous than those of the other Virtual Organisations supported since it is non driven by a specific research topic.
- The **Arts-Humanities** Virtual Organisation has been created to serve the needs of group of individuals and/or institutions in the context of the Arts and Humanities scenarios mainly focusing on the study of ancient documents.
- The **Environmental Monitoring** Virtual Organisation has been created to serve the needs of the Environmental Monitoring community. The activities carried on in monitoring the environment require extremely sophisticated means, including time and space instrumentation, analysis tools, integration and correlation of different data sources, reasoning, information and knowledge management.

VO	VREs	Users	Collections
FARM	3	37	20
Ecosystem	2	29	30
Arts-Humanities	1	21	3
EM[2]	2	N/A	27

Table 5 - D4Science-II VOs detailed information

[1] Before January 2010 the VOs and VREs serving the D4Science-II communities were operated in the context of the D4Science infrastructure and documented in [4].

[2] Dismissed in March 2010.

Eight Virtual Research Environments have been deployed and operated, each in the context of a specific Virtual Organisation:

- 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 be output in multiple formats.
- 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 **TryIt** Virtual Research Environment is conceived to serve demonstration and training activities. It gives access to a set of sample content and supports the most common functionality a gCube-based VRE might be able to support, e.g. search, browse and personal work space management.
- The **gMan** Virtual Research Environment is conceived for the activities of the homonymous project. It serves as an investigation into the support that gCube can offer to research activities that originate in the Humanities, particularly in the study of ancient documents.
- The **Environmental Monitoring (EM)** Virtual Research Environment is for making available various satellite data and services for consuming such data. It resulted from the fusion of GCM and GVM (see below).
- The **Global Ocean Chlorophyll Monitoring (GCM)** Virtual Research Environment integrates heterogeneous satellite data of microscopic marine plants and sea surface temperature maps with a pool of different sources of information related to Earth Science data and products.[3]
- The **Global Vegetation Monitoring (GVM)** Virtual Research Environment integrates heterogeneous satellite data into vegetation maps using a pool of different sources of information related to Earth Science data and products.[4]

VRE	Users	Collections
AquaMaps	33	5
FCPPS	29	16
ICIS	23	0
TryIt	27	25
gMan	18	3
EM	9	24
GCM	N/A	22
GVM	N/A	26

Table 6 - D4Science-II VREs detailed information

[3] GCM has been supported until March 2010.
[4] GVM has been supported until March 2010.

REFERENCES

[1] Ellenbroek, A. "Communities Practices and Requirements". DNA5.1 D4Science-II Project Deliverable, March 2010

[2] Candela, L.; Pagano, P. "VOs and VREs Planning". DSA2.1 D4Science-II Project Deliverable, February 2010

[3] Koltsida, P.; Kakaletris, G.; Candela, L. "VOs and VREs Operation Activity Report". DSA2.2 D4Science-II Project Deliverable, September 2010

[4] Andrade, P.; Candela, L.; Manzi, A.; Pagano P. "Infrastructure Operation Report". DSA1.3c D4Science Project Deliverable, January 2010

[5] D4Science website: http://www.d4science.eu

[6] D4Science production infrastructure wiki: https://service.wiki.d4science-ii.research-infrastructures.eu/service/index.php/SA1_Home