



**d4SCIENCE**

Project acronym	D4Science
Project full title	Distributed colLaboratories Infrastructure on Grid Enabled Technology 4 Science
Project No	212488

**Deliverable No  
DSA1.3b**

## Infrastructure Operation Report

January 2009

SEVENTH FRAMEWORK PROGRAMME  
Research Infrastructures

INFRA-2007-1.2.2: Deployment of  
e-Infrastructures for scientific communities



e-infrastructure

## DOCUMENT INFORMATION

---

Project	
Project acronym:	D4Science
Project full title:	<b>DI</b> stributed <b>coL</b> aboratories <b>I</b> nfrastructure on <b>G</b> rid <b>EN</b> abled <b>T</b> echnology <b>4</b> <b>S</b> cience
Project start:	1 <sup>st</sup> January 2008
Project duration:	24 months
Call:	INFRA-2007-1.2.2: Deployment of e-Infrastructures for scientific communities
Grant agreement no.:	212488
Document	
Deliverable number:	DSA1.3b
Deliverable title:	Infrastructure Operation Report
Contractual Date of Delivery:	January 2009
Actual Date of Delivery:	24 March 2009
Editor(s):	Pedro Andrade
Author(s):	Pedro Andrade, Pasquale Pagano
Reviewer(s):	Andrea Manieri
Participant(s):	CERN, CNR
Work package no.:	SA1
Work package title:	Infrastructure Operation
Work package leader:	CERN
Work package participants:	CNR, NKUA, UNIBASEL, ESA, FAO, WorldFish Center
Est. Person-months:	6
Distribution:	Public
Nature:	Report
Version/Revision:	1.0
Draft/Final	Final
Total number of pages: (including cover)	41
Keywords:	Production Infrastructure, Portal, VREs, Status Report

## CHANGE LOG

---

Reason for change	Issue	Revision	Date
Initial table of contents	1	0.1	Jan 09
Version for internal review	1	0.2	Feb 09
Section 2 update	1	0.3	Feb 09
Version for official review	1	1.0	Feb 09
Version after official review	1	1.1	Mar 09
Version approved by internal reviewer	1	1.2	Mar 09

## CHANGE RECORD

---

Issue: 1 Revision: 0.1

Reason for change	Page(s)	Paragraph(s)
Initial table of contents	All	All

Issue: 1 Revision: 0.2

Reason for change	Page(s)	Paragraph(s)
Version for internal review	All	All

Issue: 1 Revision: 0.3

Reason for change	Page(s)	Paragraph(s)
Section 2 update	17-28	All

Issue: 1 Revision: 1.0

Reason for change	Page(s)	Paragraph(s)
Version for official review	All	All

Issue: 1 Revision: 1.1

Reason for change	Page(s)	Paragraph(s)
Version after official review	All	All

## DISCLAIMER

---

This document contains description of the D4Science 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.research-infrastructures.eu](mailto:info@d4science.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 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 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 is a project partially funded by the European Union**

## LIST OF ABBREVIATIONS

---

CL	Collective Layer
D4Science	DIstributed colLaboratories Infrastructure on Grid Enabled Technology 4 Science
DHN	DILIGENT Hosting Node
DoW	Description of Work
EC	European Commission
EM	Environmental Monitoring
FARM	Fishery and Aquaculture Resource Management
gHN	gCube Hosting Node
IV	Infrastructure Viewer
JRA	Joint Research Activity
NA	Network Activity
RI	Running Instance
ROC	Regional Operations Centres
SA	Service Activity
TCOM	Technical Committee
VO	Virtual Organisation
VRE	Virtual Research Environment
WSRF	Web Services Resource Framework

## TABLE OF CONTENTS

---

DOCUMENT INFORMATION .....	2
CHANGE LOG .....	3
CHANGE RECORD .....	4
DISCLAIMER .....	5
LIST OF ABBREVIATIONS.....	6
TABLE OF CONTENTS .....	7
LIST OF TABLES.....	9
LIST OF FIGURES .....	10
SUMMARY .....	11
EXECUTIVE SUMMARY .....	12
1 NODES OPERATION .....	13
1.1 Introduction .....	13
1.2 Resources Allocated .....	13
1.3 Deployment and Upgrade.....	15
1.4 Certification .....	17
1.5 Downtimes.....	18
2 VIRTUAL ORGANISATIONS OPERATION .....	19
2.1 Introduction .....	19
2.2 Procedures.....	20
2.2.1 VO Registration in VOMS.....	20
2.2.2 VO Set-up.....	20
2.2.3 Deployment of VO Specific Services.....	20
2.2.4 Resources Registration .....	21
2.3 EM VO.....	21
2.3.1 VO Registration and Set-up .....	22
2.3.2 VO Specific Services.....	23
2.3.3 VO Resources.....	23
2.4 FCPPS VO .....	27
2.4.1 VO Registration and Set-up .....	27
2.4.2 VO Specific Services.....	27
2.4.3 VO Resources.....	28
2.5 ICIS VO.....	31
2.5.1 VO Registration and Set-up .....	31
2.5.2 VO Specific Services.....	31
2.5.3 VO Resources.....	32
3 INFRASTRUCTURE MONITORING .....	34
3.1 Introduction .....	34
3.2 gCube Live Monitoring .....	34
3.3 gCube Infrastructure Viewer .....	35
4 INFRASTRUCTURE SUPPORT .....	38
4.1 Incident Management .....	38
5 CONCLUSIONS .....	40

REFERENCES ..... 41



## LIST OF TABLES

---

Table 1 - Current hardware resources .....	14
Table 2 - Planned hardware resources.....	15
Table 4 - gCube nodes distribution .....	15
Table 5 - gCube gHN containers .....	16
Table 3 - gLite services distribution .....	17
Table 6 - gLite and gCube nodes downtimes .....	18
Table 7 - EM VO Services .....	23
Table 8 - EM Data Sources .....	25
Table 9 - EM Resources.....	27
Table 10 - FCPPS VO Services.....	28
Table 11 - FCPPS Data Sources.....	30
Table 12 - FCPPS Resources .....	31
Table 13 - ICIS VO Services .....	32
Table 14 - ICIS Data Sources .....	32
Table 15 - ICIS Resources.....	33
Table 16 - Incidents with gCube nodes.....	38

## LIST OF FIGURES

---

Figure 1 - Monitoring interface .....	34
Figure 2 - Infrastructure viewer gHN interface.....	35
Figure 3 - Infrastructure viewer services interface .....	36
Figure 4 - Infrastructure viewer data interface .....	36
Figure 5 - Infrastructure viewer details interface .....	37

## SUMMARY

---

This deliverable reports on the activities carried out by the D4Science Service Activity (SA) to deploy the project production infrastructure. This infrastructure can be defined as the set of hardware, software, and procedures deployed to provide D4Science user communities with a reliable production service.

In particular this document provides detailed information about the work carried out during the first year of the project to deploy and maintain the production infrastructure. This includes (1) the description of the nodes allocated to the infrastructure, (2) the presentation of the procedures defined to deploy, certify, monitor, and support the infrastructure as well as the results of their implementation, and finally (3) the report on the deployment of the project user communities VOs.

## EXECUTIVE SUMMARY

---

The objective of the D4Science Service Activity is to make available and maintain a stable production infrastructure to support the activities of the project's two user communities: Environmental Monitoring (EM) and Fishery and Aquaculture Resource Management (FARM). These communities have access to the infrastructure via the exploitation of Virtual Research Environments (VREs) that satisfy concrete needs of the communities' Virtual Organizations (VOs).

The infrastructure is based on the gCube system delivered by the project SA3 work-package, and on the gLite middleware released by the EGEE project. The operation and management of the infrastructure is facilitated by the definition and implementation of clear procedures for monitoring and support.

As originally defined in deliverable DSA1.1a [2] the production infrastructure gathers nodes provided by different partners of the D4Science consortium. The number of partners involved and the number of nodes allocated should gradually increase during the duration of the project in order to sustain the gradual increase in the provision of different VREs. This evolution is mapped to three major milestones:

- MSA1.1 due on June 2008 (M6), supporting the EM community
- MSA1.2 due on November 2008 (M11), supporting the EM and FARM communities
- MSA1.3 due on November 2009 (M23), supporting the EM and FARM communities

This report documents the work done on the production infrastructure during the first year of the project. It refers the achievement of MSA1.1, which was already reported in DSA1.3a, and describes in detail the work towards the achievement of MSA1.2.

It's not within the scope of the document to report on the infrastructure exploitation, but to describe the activities carried out (1) to maintain the infrastructure deployed in M6 (MSA1.1) up and running, and (2) to deploy the new version of the infrastructure (MSA1.2).

The deliverable is structured in five sections. Section 1 reports on the work done to manage the nodes dedicated to the infrastructure. Section 2 describes the VOs deployed in the infrastructure. Sections 3 and 4 present the activities related to the definition and implementation of procedures for monitoring and supporting the infrastructure. Finally, section 5 summarizes the deliverable by highlighting the most important achievements and difficulties reported in the deliverable.

# 1 NODES OPERATION

This section describes the hardware resources that compose the D4Science production infrastructure and the activities carried out during the first year of the project to operate such nodes. These activities include the deployment and upgrade(s) of the gHN container and gLite services on the nodes, the certification of the nodes, and the management of the nodes downtimes.

## 1.1 Introduction

The final objective of the D4Science production infrastructure is to provide Virtual Research Environments to the project user communities. These VREs are designed and deployed to serve the needs of distributed Virtual Organizations. The user communities involved in D4Science may be represented by one or more VOs. To support the provision of such diverse environments the gLite and gCube nodes play a fundamental role:

- **gCube Nodes:** gCube nodes are the hardware resources able to run gCube services. gCube is the software initially developed by the DILIGENT project now being consolidated in D4Science. The gCube software is composed by a special web service container, the gCube Hosting Node (gHN), and a set of services and libraries that provide the functionality to create, manage and exploit Virtual Research Environments (VREs);
- **gLite Nodes:** 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.

From the SA1 workpackage point of view, the operation of gLite and gCube nodes comprehends (1) the management of the physical machines that host the nodes, and (2) the tasks executed directly on the nodes to maintain them operational. Community nodes are described in section 2 since they relate to the data collections and tools exploited by the project user communities.

## 1.2 Resources Allocated

The hardware resources that composed the production infrastructure during the first year of the project (this includes MSA1.1 and MSA1.2) have been provided by 4 sites from different European countries:

- CNR: Pisa, Italy
- ESA: Rome, Italy
- NKUA: Athens, Greece
- UNIBASEL: Basel, Switzerland

The original plans also foresaw the involvement of two other sites, from FAO and WorldFish (representing the FARM community) in the production infrastructure for MSA1.2. However it was decided to first make available the FARM VREs defined for MSA1.2 and only afterwards to start exploiting the hardware resources provided by these two partners. This gives the community some time to explore their VREs and get more deeply involved. In any case preliminary conversations with FAO site administrators have already started to explain network and hardware requirements.

In deliverable DSA1.3a, which reported the status of the infrastructure after the achievement of milestone MSA1.1 (June 2008), the total number of processor cores dedicated to the infrastructure at the end of June 2008 was 69.

During the second half of 2008, a considerable effort has been invested to increase the number of nodes allocated to the infrastructure to better cope with the requirements imposed by MSA1.2. As a result of this effort, the total number of nodes fully dedicated to the production infrastructure at the end of December 2008 was 75.

Table 1 provides further information about the characteristics of these machines and how they are distributed between the different sites involved.

	#	Description	Total RAM	Total Storage	Total Cores
<b>CNR</b>	6	Sun v20z: 2x AMD Opteron(tm) Processor 252, 2.5 GHz, 1M cache	48 GB	3 TB	12
	4	Intel(R) Core(TM)2 Quad CPU Q6600 @ 2.40GHz, 4M cache	32 GB	4 TB	16
	2	AMD Opteron(tm) Processor 242, 1.6 GHz, 1M cache	16 GB	600 GB	2
<b>ESA</b>	7	Intel(R) Xeon(TM) CPU 3.20GH	16 GB	574 GB	14
	1	Intel(R) Xeon(TM) CPU 3.20GHz	4 GB	340 GB	2
	2	Intel(R) Xeon(TM) CPU 2.80GHz	4 GB	10 TB	4
<b>NKUA</b>	7	Dual Core Intel @ 2,40 GHz	56 GB	4,06 TB	14
	1	Intel(R) Xeon(R) CPU X3220 @ 2.40GHz	8 GB	500 GB	4
<b>UNIBASEL</b>	2	Single Core @ 1.8 GHz	4 GB	2,3 TB	2
	3	Single Core @ 2.4GHz	6 GB		3
	1	Dual Core @ 2.8 GHz	3 GB		2
<b>Total</b>	<b>36</b>		<b>197 GB</b>	<b>25,4 TB</b>	<b>75</b>

**Table 1 - Current hardware resources**

As the numbers above show, the evolution of processor cores from MSA1.1 to MSA1.2 is of approximately 10%. Together with the increase of processing power also the storage capacity increased to a current total of approximately 25 TB available.

It should be highlighted the high performance of most of the machines reported above. This capacity allows to take advantage of virtualization techniques, making possible the execution of more than one gLite or gCube nodes in the same physical host.

The increase of the infrastructure size is planned to continue during the second year of the project towards the final infrastructure milestone (MSA1.3). As detailed in Table 2 more 9 machines (representing 36 cores) are already planned for 2009. Others may be added at a later stage. With this addition a total of 45 machines be will at the disposal of the project. More detailed information about the resources planned for 2009 can be found in DSA1.1b [3].

	#	Description	Total RAM	Total Storage	Total Cores
<b>CNR</b>	2	Sun Fire X4100 M2 x64 Server: 2x AMD Opteron, (2.3GHz/1MB)	64 GB	584 GB	16

quad-core processor					
	2	Sun Fire X2100 M2 Opteron dual (2,8 GHz 1M L2)	16 GB	500 GB	4
	4	Sun v20z: 2x AMD Opteron(tm) Processor 252, 2.5 GHz, 1M cache	32 GB	2 TB	8
<b>FAO</b>	1	HP 2 Quad-Core Intel® Xeon® Processors X5450 @ 3.00 GHz	8 GB	292 GB	8
<b>Total</b>	<b>9</b>	<b>-</b>	<b>120 GB</b>	<b>3,4 TB</b>	<b>36</b>

Table 2 - Planned hardware resources

### 1.3 Deployment and Upgrade

As introduced before, the resources allocated to the D4Science production infrastructure have been exploited to host gCube and gLite nodes. Due to the different nature of these two types of nodes, the description of the deployment and upgrade of these is also organized in two parts.

#### gCube Nodes

As described in DSA1.3a [6] the gCube nodes have been originally deployed using the previous version of the gCube system. The version resulting from the DILIGENT project was based on the DILIGENT Hosting Node (DHN). This container was the base of the gCube nodes used in MSA1.1.

The plan towards MSA1.2 defined as main requirement the upgrade of the nodes to the new version of the gCube system, in order to reflect the new developments introduced by the project JRA activity. One main change resulting from such new development was the replacement of the old DHN container by the new gCube Hosting Nodes (gHN) container. Further details about the gHN can be found in the DJRA1.1a [7].

In order to maintain the VREs made available by MSA1.1, the transition from DHN-based nodes to gHN-based nodes was done in parallel. The old DHN nodes were kept and new nodes have been prepared to host the gHN containers. The following table report the distribution of these nodes among SA1 partners. The portal is represented independently since it does require a DHN or gHN to run.

	DHN	gHN	Portal
<b>CNR</b>	✓	✓	✓
<b>ESA</b>		✓	
<b>NKUA</b>	✓	✓	
<b>UNIBASEL</b>	✓	✓	

Table 3 - gCube nodes distribution

The number of DHN-based nodes deployed for MSA1.1 has not changed since this milestone was announced. More information about these nodes can be found in DSA1.3a. These nodes will be redeployed with the gHN container as soon as the VREs created in MSA1.1 are completely replaced by the new VREs deployed in MSA1.2.

As presented in Table 4 four sites provided gHN-based nodes. Each gHN node is exploited to host other gCube services. These services can provide functionality for the overall management of the infrastructure (enabling services) or provide VO-level functionality. In this second case they maybe configured to support one or more VOs. The following table reports the total number of gHN-based nodes provided by each site as well as which type of gCube service is supported by each site.

	Number	Enabling Services	VO Services
<b>CNR</b>	16	Yes	EM FCPPS ICIS
<b>ESA</b>	2	No	EM
<b>NKUA</b>	6	No	EM FCPPS ICIS
<b>UNIBASEL</b>	5	No	EM FCPPS ICIS
<b>TOTAL</b>	<b>29</b>		

**Table 4 - gCube gHN containers**

Since the initial deployment, the number of gHN nodes provided by each site and the support of each site to the different VOs did not change considerably.

The only changes to report are the upgrades of the gHNs to newer version. It should be highlighted that this upgrade can be executed locally by the site managers or remotely by the infrastructure administrator. In the last case, the upgrade is done by exploring the remote deployment functionality provided by the gCube VRE Manager service.

The following upgrades have been scheduled and executed in the last months:

- Initial deployment:
  - Version: 1-1-0 (from org.gcube.1-0-3)
  - Date: 22/10/2008
  - Done by: locally by site managers
- 1<sup>st</sup> Upgrade:
  - Version: 1-1-1 (from org.gcube.1-0-4/org.gcube.1-1-0)
  - Date: 10/12/2008
  - Done by: locally by site managers
- 2<sup>st</sup> Upgrade:
  - Version: 1-1-3 (from org.gcube.1-1-2)
  - Date: 21/01/2009
  - Done by: locally by site managers
- 3<sup>rd</sup> Upgrade:
  - Version: 1-1-3p (from org.gcube.1-1-2)
  - Date: 26/01/2009
  - Done by: remotely by infrastructure administrator
- 4<sup>th</sup> Upgrade (partial):
  - Version: 1-1-4 (from org.gcube.1-1-2)
  - Date: 29/01/2009
  - Done by: locally by site managers
- 5<sup>th</sup> Upgrade (partial):
  - Version: 1-1-4p (from org.gcube.1-1-4)
  - Date: 09/02/2009



- o Done by: remotely by infrastructure administrator

Before the initial deployment, two previous versions of the gHN (1-0-0 and 1-0-1) have also been deployed as an experimental activity only by some sites.

Concerning the deployment of the infrastructure portal, such activity was executed as planned. A first version of the portal was deployed for MSA1.1. It provided access to the VO/VRE management functionalities to VO/VRE administrators and VRE access to user communities' end-users. The portal is accessible at <http://portal.d4science.research-infrastructures.eu>.

Due the co-existence of DHN-based nodes for MSA1.1 and gHN-based nodes for MSA1.2, it was required to deploy a new instance of the portal for MSA1.2. The portal was deployed with the admin and user portlets included in release org.gcube.1-1-4. This version of the portal is accessible at <http://newportal.d4science.research-infrastructures.eu>. As soon as MSA1.1 VREs are replaced by the MSA1.2 VREs the previous portal will be discontinued.

### gLite Nodes

With respect to gLite nodes there are few differences to highlight when comparing with the information reported in July on DSA1.3a [6]. The table below presents how the different gLite services have been distributed between the infrastructure sites for MSA1.2.

	CE	WN	sBDII	SE	WMS	LFC	VOMS	MyProxy
<b>CNR</b>	✓	✓	✓	✓	✓		✓	✓
<b>ESA</b>	✓	✓	✓	✓		✓	✓	
<b>NKUA</b>	✓	✓	✓	✓	✓			

**Table 5 - gLite services distribution**

The services marked with a green symbol (✓) represent a primary instance and with orange symbol (✓) a secondary instance.

Comparing with the status on MSA1.1, the following changes should be mentioned:

- UNIBASEL site stop providing gLite nodes to the infrastructure. This decision was based on the need to allocate more machines from UNIBASEL as gCube nodes;
- The gLite services, VOMS and MyProxy, temporarily running at ENG during the first part of 2008, were transferred to CNR (as originally planned);
- ESA deployed a secondary instance of VOMS.

In the last months few major developments have been added to the gLite middleware by the EGEE project. The current version is still gLite 3.1. As a consequence only patches to this release have been provided. All patches were applied on the gLite nodes of the D4Science infrastructure.

## 1.4 Certification

The certification procedure for the infrastructure nodes is described in the D4Science production infrastructure website [1] under the "Certification" page.

### gCube Nodes

Concerning the gCube nodes, there are no certification problems to report. All new nodes allocated to the infrastructure to host the gHN container have been kept up and running and registered in the gCube Information System with the correct activation status. With no certification problems all gHN containers were available to host any service of the gCube system (enabling service, VO-level service, VRE-level service).

### gLite Nodes

As reported in DSA1.3a [6] all nodes running gLite services have been certified by the EGEE Regional Operation Centers during the first half of 2008. The three D4Science sites running gLite services (CNR, ESA, and NKUA) have kept their services up and running inline with the service availability levels required by EGEE, therefore no new certification process was needed.

## 1.5 Downtimes

With the VREs deployed for MSA1.1 up and running, more attention was given to the management of nodes downtimes. Again, different procedures for downtime declaration apply for gLite and gCube nodes. During the first year of the project the following downtimes have been declared:

	gCube	gLite
<b>CNR</b>	2	4
<b>ESA</b>	2	4
<b>NKUA</b>	2	11
<b>UNIBASEL</b>	1	-
<b>TOTAL</b>	<b>7</b>	<b>19</b>

**Table 6 - gLite and gCube nodes downtimes**

Due to the different procedures to announce downtimes for gLite and gCube nodes, it may happen that a common problem affecting the gLite and gCube nodes of one site is communicated in two downtime announcements. Therefore the total number of problems causing downtime situations is lower than 26 (7+19).

Most of the downtimes did not last more than 1 day so the production infrastructure has never affected by long unavailability periods. In any case, further work should be carried out to reduce the total number of downtimes and speed up the restart of the nodes from downtime status.

## 2 VIRTUAL ORGANISATIONS OPERATION

This section presents the Virtual Organisations that have been deployed in the D4Science production infrastructure to support the two user communities represented in the project. It explains the procedures defined to manage such VOs and provides detailed information about the deployment of the different VOs as well as their registered resources.

### 2.1 Introduction

The D4Science production infrastructure manages resources belonging to the VOs identified by the two large user communities participating to the D4Science project. These communities, namely Environmental Monitoring (EM) and Fisheries and Aquaculture Resources Management (FARM), identified the needs for the creation of three VOs: 1 VO for EM and 2 VOs for FARM. These different organization is motivated by the fact that the FARM community gathers people from different areas of work with different work objectives which are better represented in two separate VOs.

The **Fishery Country Profiles Production System (FCPPS)** VO currently collects resources provided by the FARM community and it is led by the FAO Fisheries and Aquaculture department. It contains the resources, i.e. services and data, needed for the generation of fisheries and aquaculture reports containing focused information in a country-specific format to enhance decision-making and promote advocacy in the sustainable use and conservation of aquatic resources. The production of country profiles requires complex aggregation and editing of continuously evolving multi-lingual data from heterogeneous sources. The FCPPS VRE organizes these resources and provides the set of features required to update these reports as frequently as the community requires. In the second phase of the project this VO plans to integrate new data sources coming from the other VOs if they will prove to be useful for the enhancement of the current environment.

The **Integrated Capture Information System (ICIS)** VO collects resources provided by the FARM community and it is led by the FAO Fisheries and Aquaculture department. It currently makes exploitable an enhanced version of the resources owned by the WorldFish Center about species distribution while it is planned to integrate the regional and global capture and distribution information of aquatic species, from a number of Regional Fishery Management Organizations (RFMOs) and the FAO into a common system. The ICIS VRE will be enhanced to support the configuration of algorithms and filters for parameters such as area, species distribution and habitat, providing an harmonized view of catch statistics and allowing the community to overlay according to pre-defined reallocation rules. Aggregation and provenance metadata will also be provided.

The **Environmental Monitoring (EM)** VO collects resources provided by the EM community. This integrated environment puts at disposal a functionality set which is not today available in Earth Science to support and perform research activities such as the process of information and the on-demand ingesting of the results; the creation of perform customized collection of information; and the aggregation of relevant information into ad-hoc information sources that are kept updated automatically. The EM VO hosts satellite data of microscopic marine plants, sea surface temperature, and vegetative land cover that represent an important source of information in many Earth system monitoring and observation processes.

The technological procedures for the creation and maintenance of such VOs are presented in section 2.1. The remaining sections report on the status, set of resources, and Virtual Research Environments characterizing each VO.

## 2.2 Procedures

The D4Science production infrastructure is able to manage one or more VOs. For each VO the following steps have to be performed.

### 2.2.1 VO Registration in VOMS

The VOs registered in the D4Science production infrastructure are modelled as VOMS groups of the `d4science.research-infrastructures.eu` VOMS VO.

For each group the following configuration parameters have to be specified:

- Default Access Control List (ACL) for the group. If not specified the ACL defined for the `d4science.research-infrastructures.eu` VOMS VO are applied;
- Membership details for the group. At least one Admin has to be identified for the group;
- Generic attributes for the group. This step is not mandatory.

These operations are performed by the D4Science production infrastructure Administrator by using the VOMS user interface accessible at:

<https://voms.research-infrastructures.eu:8443/voms/d4science.research-infrastructures.eu>.

### 2.2.2 VO Set-up

The initial step for the set-up of a new VO is the manual deployment of the gCube enabling services in two gHNs of the production infrastructure:

- The first gHN hosts the Information System consisting of the gCube IS-Registry and gCube IS-Notifier services;
- The second gHN hosts the VRE Management system consisting of the gCube VRE-Manager and gCube VRE-Modeler services.

These two subsystems are VO specific and allow to perform all other steps of the procedure till the creation of VREs by using the D4Science portal user interface.

Detailed description of how to deploy these subsystems is part of the Administration Guide accessible at:

[https://technical.wiki.d4science.research-infrastructures.eu/documentation/index.php/Administrator%27s\\_Guide](https://technical.wiki.d4science.research-infrastructures.eu/documentation/index.php/Administrator%27s_Guide)

### 2.2.3 Deployment of VO Specific Services

The deployment of the VO enabling services is the precondition for the creation of the VO environment. After their deployment, the VO Administrator can complete the creation of the VO by choosing the right configuration for the community he/she represents by using the D4Science production portal accessible at:

<http://newportal.d4science.research-infrastructures.eu>

Three subsystems must be deployed:

- gCube Content Management subsystem. It consists of the Storage Management, Content Management, and Collection Management services;
- gCube Metadata Management subsystem. It consists of the Metadata Manager, Metadata Broker, and XML Indexer services;
- gCube Index Management subsystem. It consists of Full Text, Forward, and Geospatial suites. Each suite consists of the Index Management, Index Updater, and Index Lookup services. The VO Administrator can deploy all suites or limits his/her choice to the most appropriate one for his/her VO.

Each subsystem can be deployed on one or more gHNs. All above listed services are VO specific (one instance per VO at most) but the Index Lookup services that can be replicated in multiple gHNs to improve the responsiveness and distribute the load over multiple servers.

Detailed description of how to configure these subsystems is part of the Administration Guide accessible at:

[https://technical.wiki.d4science.research-infrastructures.eu/documentation/index.php/Administrator%27s\\_Guide](https://technical.wiki.d4science.research-infrastructures.eu/documentation/index.php/Administrator%27s_Guide)

#### 2.2.4 Resources Registration

The deployment of the VO specific services completes the preparation of the VO environment that becomes ready to host data sets and resources specific of that VO.

The following resources can be registered for the VO by using the D4Science portal:

- **Data Sources.** Any data sets accessible by using a large set of protocols (HTTP, HTTPS, GridFTP, FTP, SFTP) can be linked or imported to make them exploitable by the communities through VREs. This operation is performed by instructing the gCube Archive Import Client suite through an import script.
- **Index Types.** The Index Type defines the index structure and manages what features the index provides. This process is similar to defining a schema or table structure for a database. The index type specifies each field that is contained in the index and, for each field, a list of attributes that describe what kind of information is stored. Fields in a feed that are not described in the corresponding index type will not be indexed (will not be present in the index), and thus it will not be possible to search these fields. A set of predefined Index Types configured for well know metadata formats can be used or personalized. Additional Index Types can be defined and registered.
- **Transformation Programs.** A Transformation Program is used to define the rules that the system must use to transform one metadata compliant to a format in another metadata object compliant with a different format. Transformed output can be a snapshot of its input or a mirror continually updated along with its original source. Individual transformation programs are treated as building blocks to synthesize more elaborate and complicated new programs. A set of predefined Transformation Program configured for well know set of transformation between standard metadata formats can be used or personalized. Additional Transformation Program can be defined and registered.
- **Default User Profile.** The Default User Profile contains information about the mandatory elements that all user's profiles must have. These elements include the presentation preferences for the visualization of the metadata objects and the presentation of search results for each registered metadata format; the default language; and other parameters that allow to personalize the VO. The Default User Profile can be personalized by each user after the authentication phase.

Detailed description on how to personalize the VO by registering the above described resources is available at:

<https://technical.wiki.d4science.research-infrastructures.eu/documentation/index.php/>

### 2.3 EM VO

The Environmental Monitoring VO collects resources provided by the EM community. These resources are used in two Virtual Research Environments, named Global Ocean Chlorophyll Monitoring (GCM) and Global Land Vegetation Monitoring (GVM).

The Global Ocean Chlorophyll Monitoring VRE plans to use satellite observations to represent the current state of the ocean. It is based on the assumption that the phytoplankton plays a similar role to terrestrial green plants in the photosynthetic process and are credited with removing as much carbon dioxide from the atmosphere as their earthbound counterparts, making it important to monitor and model plankton into calculations of future climate change. The phytoplankton also forms the basis of the ocean's food chain and has an essential role in the global ecosystem: it could help to

reduce the amount of carbon dioxide in the atmosphere, and are a useful indicator of changes in ocean productivity. Sea Surface Temperature (SST): SST techniques are used for monitoring and forecasting the ocean and managing the marine environment. In particular, SST measurements are utilized for operational tasks, such as providing forecasts of sea state and currents, for monitoring algal blooms or marine pollution, and predicting the fate of oil spills. There are two sensor types used to measure SST from space: infra-red radiometers and microwave radiometers.

The EM VO decided to deploy in this VRE the following set of features:

- textual and geographic search, collection inspection, a.k.a. browse;
- text, image and generic annotation support;
- personal workspace management;
- shared workspace management;
- support for the sharing of personal objects;
- report generation;

In the second part of the project the following capabilities are planned to be integrated:

- process design;
- process execution on-demand;
- ESA gPod service integration;

The Global Land Vegetation Monitoring VRE plans to consume satellite observations about vegetative land cover that represents an important variable in many Earth system processes. Many general circulation and carbon exchange models require vegetative cover as a boundary layer necessary for running the model. Vegetation also represents an important natural resource for humans and other species, and quantifying the types and extent of vegetation is important to resource management and issues regarding land cover change. To measure and map the density of green vegetation across the Earth's landscapes, satellite data are used to observe the distinct wavelengths of visible and near-infrared sunlight that is absorbed and reflected by the plants. Calculating the ratio of the visible and near-infrared light reflected back up to the sensor yields a number from minus one (-1) to plus one (+1). The result of this calculation is called the Normalized Difference Vegetation Index, or NDVI. An NDVI value of zero means no green vegetation and close to +1 (0.8 - 0.9) indicates the highest possible density of green leaves.

The EM VO decided to deploy in this VRE the following set of features:

- Textual and geographic search, collection inspection, a.k.a. browse;
- Text, image and generic annotation support;
- Personal workspace management;
- Shared workspace management;
- Support for the sharing of personal objects;
- Report generation.

In the second part of the project the following capabilities are planned to be integrated:

- Process design;
- Process execution on-demand;
- ESA gPod service integration.

### 2.3.1 VO Registration and Set-up

The Environmental Monitoring VO registration and set-up was performed by the D4Science production infrastructure administrator as detailed below:

- The `/d4science.research-infrastructures.eu/EM` group was created in VOMS. No VO specific ACL or properties were defined;

- The gHN running on node4.p.d4science.research-infrastructures.eu was selected to host the EM Information System;
- The gHN running on node3.p.d4science.research-infrastructures.eu was selected to host the EM VRE Management system.

### 2.3.2 VO Specific Services

The Environmental Monitoring VO specific services have been allocated in accordance with the following deployment diagram:

Site	gHN	Services
<b>CNR</b>	node13.p.d4science.research-infrastructures.eu	GeoIndexLookup, ForwardIndexLookup, FullTextIndexLookup
<b>CNR</b>	node12.p.d4science.research-infrastructures.eu	MetadataBroker, MetadataManager, XMLIndexer
<b>ESA</b>	grid-eo-store06.esrin.esa.int	GeoIndexLookup, ForwardIndexLookup, FullTextIndexLookup
<b>NKUA</b>	dl28.di.uoa.gr	FullTextIndexUpdater, FullTextIndexManagement
<b>NKUA</b>	dl24.di.uoa.gr	ForwardIndexManagement, GeoIndexManagement, ForwardIndexUpdater GeoIndexUpdater
<b>UNIBASEL</b>	dil04.cs.unibas.ch	CollectionManagementService, ContentManagementService, StorageManagementService

Table 7 - EM VO Services

### 2.3.3 VO Resources

The data sources described in Table 8 currently populate the EM VO. Additional details about them are reported in the Environmental Monitoring Resources Wiki page [8].

EM Data Sources
<p><b>ESA special issues Living Planet Programme and Earth Observation Handbook</b></p> <p>This collection consists of reports on the needs of the scientific communities from a social trend point of view. These reports give information about the future policies and trends of the European Space Agency related to a global scale environment and space missions analysis. These reports are either pdf documents or web pages.</p> <p>To make them available in the production infrastructure, the SA2 team created a script that instruct the Archive Import Client with <i>(/)</i> the URLs to download the pdf files or reference the web pages and <i>(/i)</i> the metadata records in Dublin Core Format.</p>
<p><b>ESA Earth Images</b></p> <p>This collection consists of Earth images collected and distributed by ESA with scientific captions, links and high-resolution formats. They are acquired from different instruments on-board of several Earth Observation satellites such as Envisat.</p> <p>To make them available in the production infrastructure, the SA2 team created a</p>



script that instruct the Archive Import Client with (i) the URLs to download the images and (ii) the metadata records a proprietary format.

### **Landsat7**

This collection consists of images covering different European cities as well as land (including forests and cropland) belonging to different European countries.

To make them available in the production infrastructure, the SA2 team performed a preliminary *processing activity* leading to the production of the images and the relative metadata. Such a processing activity consisted in fetching the initial products from ESA Storage Elements, combining them through an algorithm that produces a *regular product image* and a *false colour image* (both in low-resolution and high-resolution). This procedure has been performed by relying on the EGEE production infrastructure. Once the data to be imported was ready, SA2 created a script that instruct the Archive Import Client to create compound objects consisting of, in addition to the metadata in a proprietary schema, a main object containing the low-resolution regular image that have as alternative representations the high-resolution regular image and the low-resolution false colour image, that in its turn has the high-resolution false colour image as alternative representation.

### **MERIS Data (Meris level 3 products) Chlorophyll in water**

Microscopic marine plants (called phytoplankton) contain chlorophyll, a green pigment they use during photosynthesis. Using satellite sensors, we can measure chlorophyll concentrations in oceans, lakes and seas to indicate the distribution and abundance of phytoplankton. Phytoplankton are the base of the marine food chain and, therefore, are a good indicator of the abundance of life in a body of water. This data source contains Chlorophyll-a, case-1 water (chl1) yearly level 3 products for the years 2002-2008. Each yearly object contains monthly products.

To make them available in the production infrastructure, the SA2 team created a script that by scavenging web pages acquires the URLs of the various products, assembles them in yearly compound objects and produces the relative metadata. Each early product consists of the monthly products each having its own metadata and multiple parts ranging from the image representing them in the available formats, the official URL of such products, and the URLs of any other kind of format the product is available.

### **MERIS Data (Meris Land level 3 products) Global Vegetation Index**

This data source contains products related to the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR). Establishing long term time series of remote sensing geophysical products, that are relevant for environmental applications at regional or global scale, calls for the production of high level products, i.e. at a lower spatial resolution relevant for modelling communities such as the carbon or climate domain. The spatially aggregated land products are therefore produced to be directly used for regional or global scale land analysis. This collection contains yearly level 3 products for the years 2002-2008. Each yearly object contains monthly products.

To make them available in the production infrastructure, the SA2 team created a script that by scavenging web pages acquires the URLs of the various products, assembles them in yearly compound objects and produces the relative metadata. Each early product consists of the monthly products each having its own metadata and multiple parts ranging from the image representing them in the available formats, the official URL of such products, and the URLs of any other kind of format the product is available, e.g. the row data.

### **European Environment Agency (EEA) Reports**

This collection is useful as data source for Environmental reports, for both Ocean and Land monitoring applications. The European Environment Agency (EEA) produce reports describing the environment's status. The EEA papers provide a wide range of



<p>information related to analyze the state of and trends in the environment together with scenario development, policy evaluation and data quality assurance. The information provided by the EEA comes from a wide range of sources. Several reports and short briefings on particular issues are published every year. Briefings and summaries of major reports are usually translated into the official languages of EEA member countries.</p> <p>To make them available in the production infrastructure, the SA2 team created a script that by scavenging the web pages EEA reports are exposed gathers the content of such reports and create compound objects for them.</p>
<p><b>Mediterranean Action Plan (MAP) Technical Reports</b></p> <p>This collection is particularly well suited for ocean monitoring applications. This reports are focused on Mediterranean countries, for assessing and controlling marine pollution. The main objectives of the MAP (Mediterranean Action Plan) are to assist the Mediterranean countries to assess and control marine pollution, to formulate their national environment policies, to improve the ability of governments to identify better options for alternative patterns of development, and to optimize the choices for allocation of resources. Recently, coastal zone planning and management has been integrated as a key tool through which solutions are planned.</p> <p>To make them available in the production infrastructure, the SA2 team created a script that by scavenging the web pages MAP reports are exposed gathers the content of such reports, automatically generates the metadata and creates compound objects for them.</p>
<p><b>Advanced Along Track Scanning Radiometer (AATSR) Data</b></p> <p>AATSR data set contains data set for Sea Surface Temperature (SST) analysis, produced by the service Medspiration SST. The Medspiration Project is a European initiative to combine data measured independently by several different satellite systems into a set of data products that represent the best measure of SST, presented in a form that can be assimilated into ocean forecasting models.</p> <p>To make them available in the production infrastructure, the SA2 team created a script that provides the Archive Import Client with the URLs of the products and the relative metadata.</p>

**Table 8 - EM Data Sources**

The VO specific resources described in Table 9 currently populate the EM VO. Additional details about them are reported in the Infrastructure Viewer accessible at:

<http://monitor.d4science.research-infrastructures.eu/iv/>

EM Resources	
<b>Index Types</b>	The EM data sources are represented with a set of metadata formats always written in English. This set includes ISO19115, Dublin Core, Earth Science (ES), and eiDB formats. Full text and forward indices were created for all of them. Geospatial indices were created for the metadata formats including geographical coordinates.
<b>Transformation Programs</b>	Transformation programs (TPs) were defined to map all collections to the pivot schema selected by the EM VO, namely the ES format, and to create the different type of indices. The following TPs were registered:  DC -> ES - This transformation program transforms a given collection adopting the DC schema to a collection

in ES schema.

eiDB -> ES - This transformation program transforms a given collection adopting the eiDB schema to a collection in ES schema.

DC -> Full Text RowsetRS - This transformation program transforms a given collection adopting the DC schema to a ResultSet in rowset format (suitable for feeding the full text search index).

DC -> Forward RowsetRS - This transformation program transforms a given collection adopting the DC schema to a set of ResultSets in forward rowset format (suitable for feeding the forward search index).

eiDB -> Geo RowsetRS - This transformation program transforms a given collection adopting the eiDB schema to a ResultSet in GeoRowset format ( suitable for feeding the geo index).

eiDBCol -> Full Text RowsetRS - This transformation program transforms a given collection adopting the eiDB schema to a ResultSet in rowset format (suitable for feeding the full text search index).

eiDBCol -> Forward RowsetRS - This transformation program transforms a given collection adopting the eiDBCol schema to a set of ResultSets in forward rowset format (suitable for feeding the forward search index).

ES -> Full Text RowsetRS - This transformation program transforms a given collection adopting the ES schema to a ResultSet in rowset format (suitable for feeding the full text search index).

ES -> Forward RowsetRS - This transformation program transforms a given collection adopting the ES schema to a set of ResultSets in forward rowset format (suitable for feeding the forward search index).

ES -> Geo RowsetRS - This transformation program transforms a given collection adopting the ES schema to a ResultSet in GeoRowset format ( suitable for feeding the geo index).

### Generic Resources

The following resources have been customised for the EM VO:

- MetadataSchemaInfo: It's needed for every schema that has been adopted by the VO. It has been defined for DC, eiDB, and ES schema. Each MetadataSchemaInfo contains the static configuration describing: searchable fields, fields that can be used to sort the results, and the search operator that can be applied on every searchable field.

- MetadataXSLT, PresentationXSLT, TitleXSLT, and BrowseFieldXSLT: default xslt files for the metadata are needed for every schema adopted by the VO in order to define the presentation of metadata objects. They have been defined for DC, eiDB, and ES schema. Users can register additional presentation xslt files to personalize the presentation layout.

**Table 9 - EM Resources****2.4 FCPPS VO**

The Fisheries Country Profiles Production System VO collects resources provided by the FARM community. These resources are used in the homonymous VRE that is dedicated to fisheries and aquaculture authors, managers and researchers who produce reports containing country-level data. The FCPPS is a Virtual Research Environment that manages access to multiple data sources divided by domain, 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. Unlike Microsoft Office our product has a workflow and structured data that enables collaboration, annotation and versioning of the underlying data sources that make up the final product. It enforces a structured process and structured data while allowing users flexibility in actual content generation.

The FCCPS VO decided to deploy in this VRE the following set of features:

- Geographical search by country, region or water area;
- Text search within text corpus (boolean, case-insensitive, phrase-matching);
- Search by metadata with field selection dynamically built from schema (boolean, case-insensitive, phrase-matching);
- Report generation including rich-text editing, drag and drop of images and text; hotspots dynamically updated based on association to underlying object.
- Annotation of objects and baskets;
- Personal workspace management;
- Shared workspace management;
- Support for the sharing of personal objects.

**2.4.1 VO Registration and Set-up**

The Fisheries Country Profiles Production System VO registration and setup was performed by the D4Science production infrastructure administrator as detailed below:

- The /d4science.research-infrastructures.eu/FCCPS group was created in VOMS. No VO specific ACL or properties were defined;
- The gHN running on node11.p.d4science.research-infrastructures.eu was selected to host the EM Information System;
- The gHN running on node10.p.d4science.research-infrastructures.eu was selected to host the EM VRE Management system.

**2.4.2 VO Specific Services**

The Fisheries Country Profiles Production System VO specific services have been allocated in accordance with the following deployment diagram:

Site	gHN	Services
<b>CNR</b>	node14.p.d4science.research-infrastructures.eu	GeoIndexLookup, ForwardIndexLookup, FullTextIndexLookup
<b>CNR</b>	node10.p.d4science.research-infrastructures.eu	MetadataBroker, MetadataManager, XMLIndexer
<b>NKUA</b>	dl25.di.uoa.gr	GeoIndexLookup, ForwardIndexLookup,

		FullTextIndexLookup
<b>NKUA</b>	dl23.di.uoa.gr	FullTextIndexUpdater, FullTextIndexManagement
<b>NKUA</b>	dl27.di.uoa.gr	ForwardIndexManagement, GeoIndexManagement, ForwardIndexUpdater GeoIndexUpdater
<b>UNIBASEL</b>	dil05.cs.unibas.ch	CollectionManagementService, ContentManagementService, StorageManagementService

**Table 10 - FCPPS VO Services****2.4.3 VO Resources**

The data sources described in Table 11 currently populate the FCPPS VO. Additional details about them are reported in the Fishery and Aquaculture Resources Wiki page [9].

FCPPS Data Sources
<p><b>Current National Aquaculture Sector Overview (NASO) Fact Sheets</b></p> <p>The National Aquaculture Sector Overview (NASO) collection consists of concise and comprehensive products, providing a general overview of the aquaculture and culture based fisheries aspects at the national level. The NASOs contain detailed information on the history of aquaculture; human resources involved in the sector; farming systems distribution and characteristics; main cultured species contributing to national production; production statistics; description of the main domestic markets and trade; promotion and management of the sector; and development trends and issues at the national level. NASO is part of the FAO Fisheries and Aquaculture Department regular programme.</p> <p>To make them available in the production infrastructure, the SA2 created an Archive Import Client script that analyses the products exposed through an RSS feed and produces compound objects representing the fact sheets (the HTML and XML version) in all the languages they exist (English, Chinese, Spanish, and French). This leads to one Content Collection and four Metadata Collections.</p>
<p><b>Current National Aquaculture Legislation Overview (NALO) Fact Sheets</b></p> <p>The National Aquaculture Legislation Overview (NALO) consist of a series of comparative national overviews of aquaculture laws and regulations from the top 40 aquaculture producing countries, and have been prepared in collaboration with the FAO Development Law Service. The preparation of the overviews is primary based on FAOLEX, a legislative database containing the world's largest electronic collection of national laws and regulations on food and agriculture, following a standard template. Further material has been collected from national sources. NALO fact sheets are updated on a regular basis, every 2/3 years.</p> <p>To make them available in the production infrastructure, the SA2 created an Archive Import Client script that analyses the products exposed through an RSS feed and produces compound objects representing the fact sheets (the HTML version with the XML version as alternative representation) in all the languages they exist (English, Chinese, Spanish, and French) with the relative metadata. This leads to one Content Collection and four Metadata Collections.</p>
<p><b>Introduced Species Fact Sheets</b></p> <p>The FAO Database on Introductions of Aquatic Species (DIAS) includes records displayed as factsheets of species introduced or transferred from one country to</p>

another. It covers freshwater, molluscs, crustaceans and marine species. This collection contains fact sheets generated from DIAS.

To make them available in the production infrastructure, the SA2 created an Archive Import Client script that analyses the products exposed through an RSS feed and produces compound objects representing the fact sheets (the HTML version with the XML version as alternative representation) with the relative metadata. This leads to one Content Collection and one Metadata Collections.

#### **Cultured Species Fact Sheets**

The Cultured Aquatic Species Fact Sheet collection is directed at those wishing to gain an understanding of the steps that should be followed to start to raise aquatic species (fish, mollusc, seaweeds and frogs) and learn about current techniques at the global level, provides a general overview of various cultural aspects (e.g. production systems, diseases and control measures, production statistics, market and trade) for the most important species in aquaculture. A detailed and comprehensive bibliography as well as related links are also provided. The fact sheets are compiled following a standard template and are regularly updated.

To make them available in the production infrastructure, the SA2 created an Archive Import Client script that analyses the products exposed through an RSS feed and produces compound objects representing the fact sheets (the HTML version with the XML version as alternative representation) in all the languages they exist (English, Chinese, Spanish, and French) with the relative metadata. This leads to one Content Collection and four Metadata Collections.

#### **Aquaculture Time Series graphs**

This data source contains graphs reporting on aquaculture statistics on production volume and value by country.

To make them available in the production infrastructure, the SA2 created an Archive Import Client script that analyses the products exposed through an RSS feed and produces objects containing the images and the relative metadata. This leads to one Content Collection and one Metadata Collections.

#### **Capture Time Series graphs**

This data source contains graphs reporting on fishery capture production statistics by country.

To make them available in the production infrastructure, the SA2 created an Archive Import Client script that analyses the products exposed through an RSS feed and produces objects containing the images and the relative metadata. This leads to one Content Collection and one Metadata Collections.

#### **Fleet Time Series graphs**

This data source contains graphs reporting national annual statistics on the number, total tonnage and total power of fishing vessels utilised in commercial, industrial and artisanal fisheries by country.

To make them available in the production infrastructure, the SA2 created an Archive Import Client script that analyses the products exposed through an RSS feed and produces objects containing the images and the relative metadata. This leads to one Content Collection and one Metadata Collections.

#### **Production Time Series graphs**

This data source contains graphs reporting on statistics on the annual production of fishery commodities and imports and exports (including re-exports) of fishery commodities by country in terms of volume and value.

To make them available in the production infrastructure, the SA2 created an Archive Import Client script that analyses the products exposed through an RSS feed and produces objects containing the images and the relative metadata. This leads to one

Content Collection and one Metadata Collections.
<p><b>Trade Time Series graphs</b></p> <p>This data source contains graphs reporting on statistics on the annual production of fishery commodities and imports and exports (including re-exports) of fishery commodities by country in terms of volume and value.</p> <p>To make them available in the production infrastructure, the SA2 created an Archive Import Client script that analyses the products exposed though an RSS feed and produces objects containing the images and the relative metadata. This leads to one Content Collection and one Metadata Collections.</p>
<p><b>Country Maps</b></p> <p>This data source consists of NOAA Satellite Image - Data from ESRI's World Worldsat Colour Shaded Relief Image. It is based on 1996 NOAA weather satellite images, with enhanced shaded relief imagery and ocean floor relief data (bathymetry) to provide a land and undersea topographic view.</p> <p>To make them available in the production infrastructure, the SA2 created an Archive Import Client script that analyses the products exposed though an RSS feed and produces objects containing the images and the relative metadata. This leads to one Content Collection and one Metadata Collections.</p>
<p><b>Species Distribution Maps</b></p> <p>This data source consists of maps showing the geographic distribution of several hundred species of major commercial interest based on habitat descriptions and geographic distributions derived from the official FAO Catalogues of Species.</p> <p>To make them available in the production infrastructure, the SA2 created an Archive Import Client script that analyses the products exposed though an RSS feed and produces objects containing the images and the relative metadata. This leads to one Content Collection and one Metadata Collections.</p>

**Table 11 - FCPPS Data Sources**

The VO specific resources described in Table 12 currently populate the FCPPS VO. Additional details about them are reported in the Infrastructure Viewer accessible at:

<http://monitor.d4science.research-infrastructures.eu/iv/>

FCPPS Resources	
<b>Index Types</b>	The FCPPS data sources are represented in a customised Dublin Core qualified metadata format written in English, Spanish, French, and Chinese. Full text and forward indices were created for all data sources.
<b>Transformation Programs</b>	<p>Transformation programs (TPs) were defined to create the different type of indices. No pivot schema has been defined since all collections are described using a common metadata schema. The following TPs were registered:</p> <p>FARM_DC -&gt; Full Text RowsetRS - This transformation program transforms a given collection adopting the customised qualified DC schema adopted by the FCPPS VO to a ResultSet in rowset format (suitable for feeding the full text search index).</p> <p>FARM_DC -&gt; Forward RowsetRS - This transformation program transforms a given collection adopting the</p>

	FARM_DC schema to a set of ResultSets in forward rowset format (suitable for feeding the forward search index).
<b>Generic Resources</b>	<p>The following resources have been customised for the FCPPS VO:</p> <ul style="list-style-type: none"> <li>- MetadataSchemaInfo: It has been defined for the FARM_DC schema. It contains the static configuration describing: searchable fields, fields that can be used to sort the results, and the search operator that can be applied on every searchable field.</li> <li>- MetadataXSLT, PresentationXSLT, TitleXSLT, and BrowseFieldXSLT: default xslt files for the FARM_DC metadata have been registered to define the presentation of FARM_DC objects. Users can register additional presentation xslt files to personalize the presentation layout.</li> </ul>

**Table 12 - FCPPS Resources**

## 2.5 ICIS VO

### 2.5.1 VO Registration and Set-up

The Integrated Capture Information System VO registration and setup was performed by the D4Science production infrastructure administrator as detailed below:

- The /d4science.research-infrastructures.eu/ICIS group was created in VOMS. No VO specific ACL or properties were defined;
- The gHN running on node9.p.d4science.research-infrastructures.eu was selected to host the EM Information System;
- The gHN running on node8.p.d4science.research-infrastructures.eu was selected to host the EM VRE Management system.

### 2.5.2 VO Specific Services

The Integrated Capture Information System VO specific services have been allocated in accordance with the following deployment diagram:

Site	gHN	Services
<b>CNR</b>	node6.p.d4science.research-infrastructures.eu	GeoIndexLookup, ForwardIndexLookup, FullTextIndexLookup
<b>CNR</b>	node5.p.d4science.research-infrastructures.eu	MetadataBroker, MetadataManager, XMLIndexer
<b>NKUA</b>	dl26.di.uoa.gr	GeoIndexLookup, ForwardIndexLookup, FullTextIndexLookup
<b>UNI BASEL</b>	dil01.cs.unibas.ch	FullTextIndexUpdater, FullTextIndexManagement, ForwardIndexManagement, GeoIndexManagement, ForwardIndexUpdater



GeoIndexUpdater		
<b>UNIBASEL</b>	dil06.cs.unibas.ch	CollectionManagementService, ContentManagementService, StorageManagementService

**Table 13 - ICIS VO Services**

**2.5.3 VO Resources**

The Data Sources described in Table 14 currently populate the FCPPS VO. Additional details about them are reported in the Fishery and Aquaculture Resources Wiki page [9].

AquaMaps
<p><b>AquaMaps</b></p> <p>This is a collection consisting of AquaMaps, i.e. predictive documents resulting in Earth maps enriched thus to show the likelihood that a certain species or a combination of species will live in specific regions or areas.</p> <p>To make them available in the production infrastructure, two complementary procedures are needed. The first one is dedicated to estimate the probabilities of species occurrence by matching the species environmental envelope, i.e. the tolerances of a species with respect to environmental parameters, against environmental conditions. The second one is dedicated to the actual production of the maps by plotting the estimated values on Earth maps. For the time being, the first activity was conducted out of the project scope. The procedure of AquaMaps production by using a data source capturing the data of more that 9'000 species and consisting of more than 50 million probability entries has been performed by the SA2 team to produce per species maps.</p> <p>The SA2 team then created an Archive Import Client script that produces compound objects representing an AquaMap for a single species and consisting in 15 images (a 2D map of the whole Earth plus 2 3D views for the Poles, 6 3D views for the Continents, and 6 3D views for the Oceans) with the relative metadata. This leads to one Content Collection and one Metadata Collections.</p>

**Table 14 - ICIS Data Sources**

The VO specific resources described in Table 15 currently populate the ICIS VO. Additional details about them are reported in the Infrastructure Viewer accessible at:

<http://monitor.d4science.research-infrastructures.eu/iv/>

ICIS Resource	
<b>Index Types</b>	The ICIS data source are represented in a 'aquamaps' metadata format written in English. Full text and forward indices were designed for all data sources.
<b>Transformation Programs</b>	Transformation programs (TPs) were defined to create the different type of indices. No pivot schema has been defined since all collections are described using a common metadata schema. The following TPs were registered:  aquamaps -> Full Text RowsetRS - This transformation program transforms a given collection adopting the customised auquamaps schema to a ResultSet in rowset format (suitable for feeding the full text search index).



	<p>aquamaps -&gt; Forward RowsetRS - This transformation program transforms a given collection adopting the aquamaps schema to a set of ResultSets in forward rowset format (suitable for feeding the forward search index).</p>
<b>Generic Resources</b>	<p>The following resources have been customised for the ICIS VO:</p> <ul style="list-style-type: none"><li>- MetadataSchemaInfo: It has been defined for the aquamaps schema. It contains the static configuration describing: searchable fields, fields that can be used to sort the results, and the search operator that can be applied on every searchable field.</li><li>- MetadataXSLT, PresentationXSLT, TitleXSLT, and BrowseFieldXSLT: default xslt files for the aquamaps metadata have been registered to define the presentation of aquamaps objects. Users can register additional presentation xslt files to personalize the presentation layout.</li></ul>

**Table 15 - ICIS Resources**

### 3 INFRASTRUCTURE MONITORING

This section introduces the monitoring tools defined and exploited during the first year of the project to monitor the status of the D4Science production infrastructure and its different resources.

#### 3.1 Introduction

Monitor a distributed infrastructure that hosts hundreds of different resources is a complex task and requires dedicated tools that simplify the identification of problems and suggest potential warnings that require further analysis.

The D4Science production infrastructure has been monitored through tools specialized for gLite and gCube resources. The complete list of tools exploited in the first year of the project is described at:

<https://infrastructure.wiki.d4science.research-infrastructures.eu/production/index.php/Monitoring>

The gCube monitor technology offers two tools to monitor the infrastructure, named gCube Live Monitoring and gCube Infrastructure Viewer (IV). These tools are accessible through web interfaces consumable also via mobile devices. This allows administrators and site managers to be always informed about the status of the infrastructure and its resources.

#### 3.2 gCube Live Monitoring

The gCube live monitoring tool allows to access resource profiles as they are stored in the gCube Information System. Colours and error messages are used to highlight the resources that are not properly exploitable by the services of the infrastructure.

In the following figures there are examples of uncertified gHN and running service instance resources.

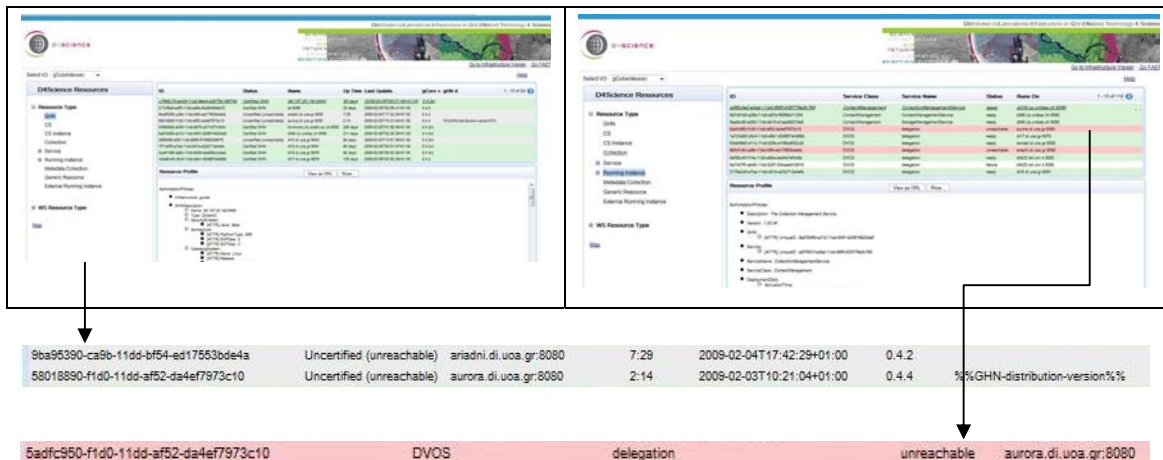


Figure 1 - Monitoring interface

For any uncertified gHN it is reported the last time the gHN worked properly and all information that allows to identify it properly. The unreachable running service instances are highlighted in red to simplify their identification.

The gCube live monitoring tool not only allows to identify resources that are failing but it allows to access each resource profile and to inspect the status of the resource, how it was configured, and all its properties.

The gCube live monitoring tool is accessible at:  
<http://monitor.d4science.research-infrastructures.eu/monitoring>

### 3.3 gCube Infrastructure Viewer

The gCube infrastructure viewer provides simplified views of the status of the infrastructure, easily consumable from any web enabled device. It exploits Treemap technology that displays hierarchical (tree-structured) resources as a set of nested rectangles. Each resource is represented as a rectangle, which is then tiled with smaller rectangles representing sub-resources. Each rectangle has an area proportional to a specified dimension on the resource. The rectangle, its border or its title are coloured to highlight additional information about the resource.

When the colour and size dimensions are correlated in some way with the tree structure, an administrator can often easily see patterns that would be difficult to spot in other ways. Moreover making an efficient use of space this tool can legibly display thousands of items on the screen simultaneously.

The gCube IV provides access to three different views:

- **GHN View** represents Sites and gHNs characterized by the RIs;
- **Services View** represents Class of Services, Services, and RIs;
- **Data View** represents Collections, Metadata Collections, and Indexes.

In the following figure there is an example of uncertified gHN resources.

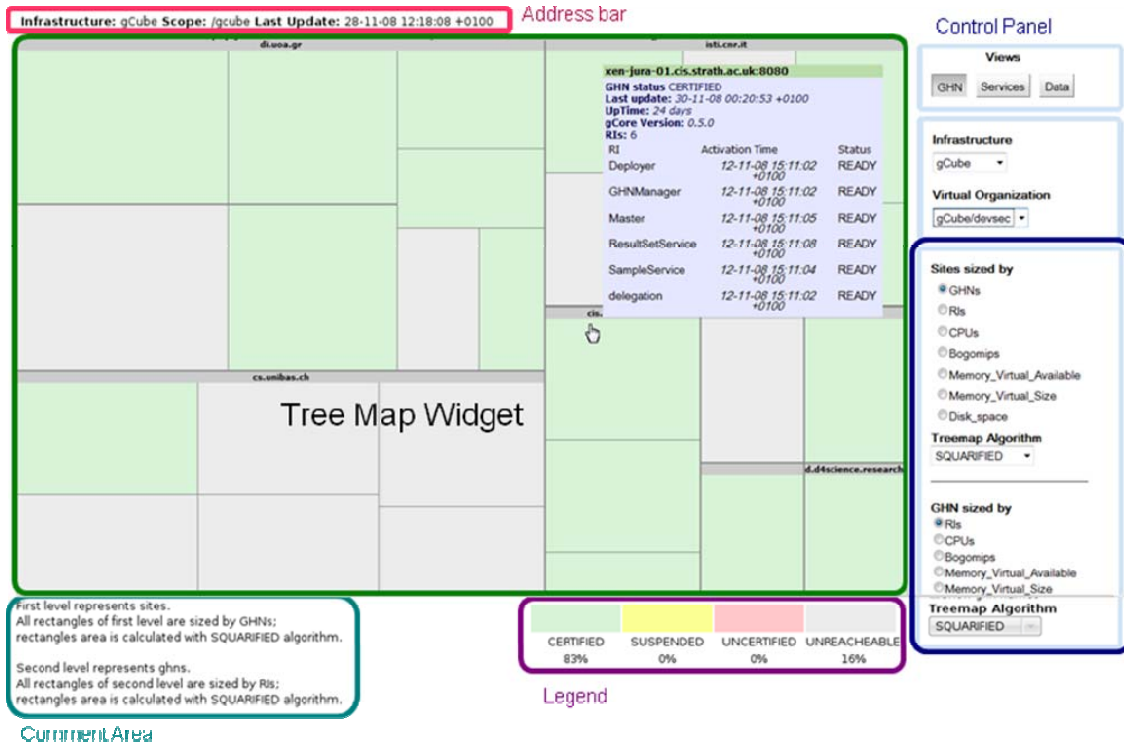


Figure 2 - Infrastructure viewer gHN interface

Different metrics and different status information are available for the different views. Figure 3 shows an example of the service view. Even without reading the test it is immediately evident that there are a number of outdated running service instances and one failed service.

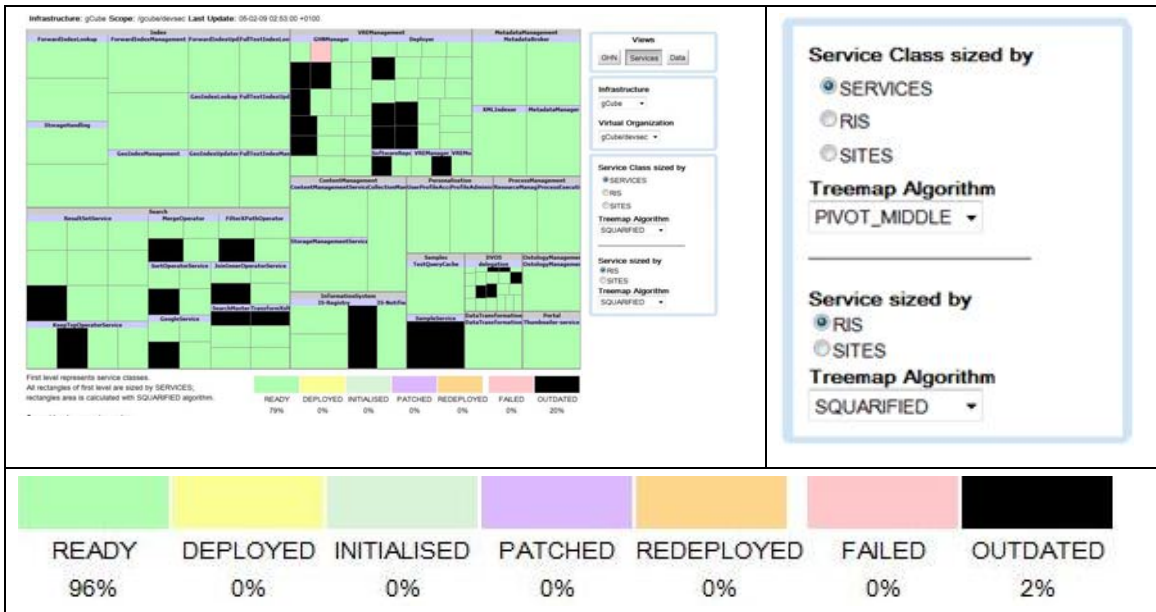


Figure 3 - Infrastructure viewer services interface

In the following figure it is shown an example of the data view. Even without reading the test it is immediately evident that there are two collections that are not accessible and one collection that should be monitored carefully.

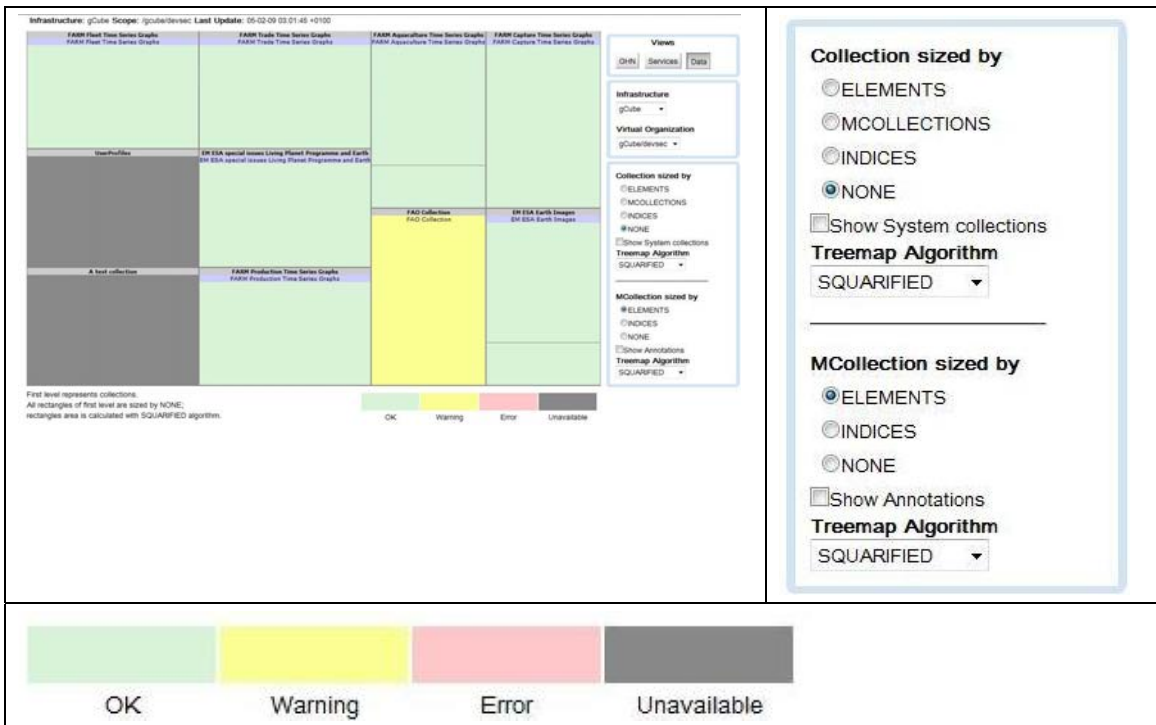
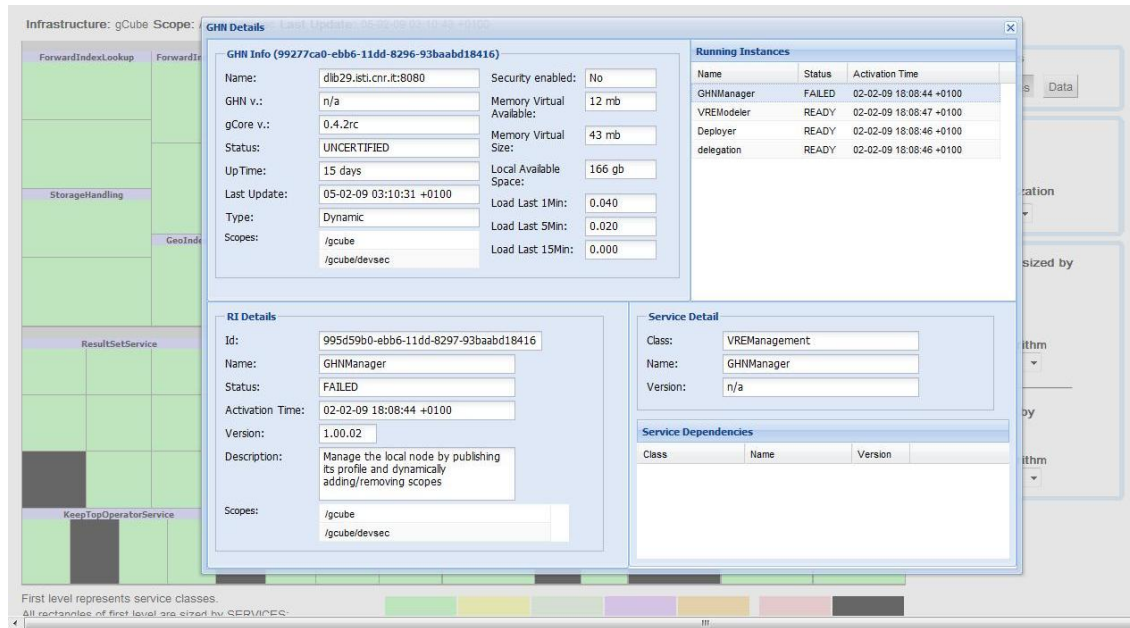


Figure 4 - Infrastructure viewer data interface

After the identification of an error, a warning, or unavailability of some resources it is possible to collect more information about the problem by clicking on the rectangle reporting the issue. A detailed panel is displayed that contains the resource specific information required to understand and correct the infrastructural problem.

In the next figure the detailed panel that appears in the Service view is shown.



**Figure 5 - Infrastructure viewer details interface**

All figures presented in this section are taken from the gCube Infrastructure Viewer (IV) running in the development infrastructure in order to better highlight how different problems are represented in the tool.

The gCube IV for the D4Science production infrastructure is accessible at:

<http://monitor.d4science.research-infrastructures.eu/iv>

## 4 INFRASTRUCTURE SUPPORT

The section describes the activities carried out to provide support to the operation and exploitation of the D4Science production infrastructure by the different infrastructure users: VO administrators, VRE administrators, Site Managers, etc. The production infrastructure support activity is based on the incident management procedure.

### 4.1 Incident Management

With the availability of the infrastructure to end-users to exploit the VREs deployed for MSA1.1, a re-structuring of the infrastructure support procedure was needed. Following the ITIL<sup>1</sup> methodology a new procedure for incident management has been defined and placed in production. This procedure replaces the previous support follow-up procedure defined in the beginning of the project.

The incident management procedure consists on the definition of different functional and hierarchical levels of support, to make sure that all incidents are correctly solved within a maximum allowed time.

An important modification introduced with the new incident management procedure was the creation of the Support Team. This team is mainly responsible for the 1<sup>st</sup> line support. In addition, it is also responsible for:

- Analyze all new incidents, categorize them, and assign a responsible person;
- If needed, escalate tickets to the TCom for 2<sup>nd</sup> level support;
- Ensure that all tickets are properly documented.

The support team is composed by people from different activities of the project (SA1, SA2, JRA4, and NA5) in order to make sure that it gathers enough knowledge to resolve any type of incident. Detailed information about the procedures, the people involved, and the workflow of the different activities, is available in the D4Science infrastructure website [1] 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.

Table 16 reports the total number of tickets submitted during the first year of the project (M1 to M12) and the average resolution time.

	Ticket Submitted	Average Duration
<b>Quarter 1</b>	0	0
<b>Quarter 2</b>	16	21,69
<b>Quarter 3</b>	53	8.94
<b>Quarter 4</b>	6	5.5
<b>Total</b>	<b>75</b>	<b>11.39</b>

**Table 16 - Incidents with gCube nodes**

<sup>1</sup> <http://www.iti1-officialsite.com/home/home.asp>

In total 79 tickets have been submitted and closed during the first year of the project. These tickets have remained open in average during 11,39 days. From the above table several conclusions can be extracted:

- The numbers of tickets opened in the different quarters reflects the work developed during the first year of the project. During Q2 several tickets have been create to manage the deployment of MSA1.1. After the release of MSA1.1 on Q3 and the subsequent usage of the deployed VREs by a larger number of users, the number of incidents has increased significantly. On the 4<sup>th</sup> quarter with the VREs usage stabilizing, the number of incidents decreased again, reflecting the small number of issues that appeared during the deployment of MSA1.2.
- When comparing Q2 with Q4, the quarters that anticipated the release of MSA1.1 and MSA1.2, it's visible a decrease of incident tickets, which is a direct consequence of the improvements introduced by the new gCube system.
- The average ticket resolution period decreased from quarter to quarter. This reduction is associated to be introduction of the Support Team that enabled a quicker resolution of the infrastructure incidents.

### **gLite Nodes**

As for gCube 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 must be submitted directly to EGEE Regional Operational Centre (ROC).

During the first six months of the project, a considerable effort was dedicated to the initial deployment of gLite. This resulted in a number of support requests from all D4Science gLite sites to the corresponding EGEE ROCs.

During the second half of 2008, with the gLite sites running gLite already deployed and no major upgrades to be executed (as explained in section 1.3) the number of support requests related to gLite deployment decreased.

Concerning the support to the exploitation of gLite nodes there are no support requests to register.



## 5 CONCLUSIONS

As planned in the project description of work, during the first year of the project, the D4Science production infrastructure has been deployed and made available to the project user communities. The first milestone of the infrastructure (MSA1.1) has been completed on July 2008 and has offered a number of VREs to the Earth Monitoring community. An addition VRE for demonstration purposes was also set-up.

Since then, the infrastructure was kept up and running and efficient support has been provided to all its users (VRE administrators, site managers, end-users, etc). In parallel to this activity a new instance of the infrastructure, running the refactored gCube system, has been deployed. This corresponds to the second infrastructure milestone (MSA1.2) and supports not only to the EM community but also to the Fisheries and Aquaculture Resources Management (FARM) community. This milestone was completed in the beginning of February 2009: the infrastructure nodes were made available, VO services deployed, content successfully imported, and appropriate user/roles assigned. New VREs can now be created and made available to the two user communities.

Apart from the mainstream progresses towards the infrastructure milestones, a number of other important achievements should also be highlighted:

- The management of infrastructure incidents was improved as a result of (1) the redesign of the support procedure based on ITIL best practices, and (2) the definition of a support team gathering people from different areas of the project to act as 1<sup>st</sup> line support;
- The deployment of the gCube nodes and in particular of the gHN container was simplified. In addition, the upgrade of these containers can now be done remotely by the infrastructure administrator without the need to contact local site managers;
- The deployment of the VO services stack for the different VOs of the infrastructure can now be executed remotely by the VO administrator based on the new gCube VRE Manager service functionality, accessible via the D4Science Portal. The deployment time of any (new) VO was therefore reduced significantly.
- A number of improvements have been introduced in the infrastructure monitoring tool. These improvements consist on the availability of more inter-connected information and a new graphical representation of the infrastructure resources.

On the other hand some differences to the original plan should also be clarified:

- The gLite nodes provided by UNIBASEL during the first months of 2008 have been discontinued. As explained, this was motivated by the limited number of resources and the need to allocate more machines as gCube Nodes in UNIBASEL.
- MSA1.2 was not delivered on November 2008 (M11) as planned. It has been achieved with approximately 2 month of delay. This delay was due to the late provision of some gCube components (services and portlets) and to the need to introduce several bug fixes on critical problems discovered during the integration and testing phases.
- Even if initial contacts have already been established with FAO site managers to start providing gCube and gLite nodes, this site is still not officially contributing to MSA1.2. This contribution was postponed for after the release of the FARM VREs.

For the second year of the project the main goal of the D4Science Service Activity is to improve the quality of service provided by the infrastructure (more efficient infrastructure support, easier management tools, and improved VO/VRE functionality) while at the same time extend the infrastructure with new hardware and data resources to support new VREs.



## REFERENCES

---

- [1] D4Science production infrastructure [ONLINE]  
<https://infrastructure.wiki.d4science.research-infrastructures.eu>
- [2] D4Science deliverable DSA1.1a “Procedures and resources plan” [ONLINE]  
<https://infrastructure.wiki.d4science.research-infrastructures.eu/production/index.php/DSA1.1a>
- [3] D4Science deliverable DSA1.1b “Procedures and resources plan” [ONLINE]  
<https://infrastructure.wiki.d4science.research-infrastructures.eu/production/index.php/DSA1.1b>
- [4] D4Science deliverable DSA1.2a “Middleware deployment and operation support procedures” [ONLINE]  
<https://infrastructure.wiki.d4science.research-infrastructures.eu/production/index.php/DSA1.2a>
- [5] D4Science deliverable DSA1.2b “Middleware deployment and operation support procedures” [ONLINE]  
<https://infrastructure.wiki.d4science.research-infrastructures.eu/production/index.php/DSA1.2b>
- [6] D4Science deliverable DSA1.3a “Infrastructure operation report”  
[http://bscw.research-infrastructures.eu/bscw/bscw.cgi/d85323/D4Science\\_DSA1.3a\\_M7.pdf](http://bscw.research-infrastructures.eu/bscw/bscw.cgi/d85323/D4Science_DSA1.3a_M7.pdf)
- [7] D4Science deliverable DJRA1.1a “D4Science high-level design”  
[http://bscw.research-infrastructures.eu/bscw/bscw.cgi/d100882/D4Science\\_DJRA1.1a\\_M6.pdf](http://bscw.research-infrastructures.eu/bscw/bscw.cgi/d100882/D4Science_DJRA1.1a_M6.pdf)
- [8] D4Science EM community resources [ONLINE]  
[https://infrastructure.wiki.d4science.research-infrastructures.eu/production/index.php/Environmental\\_Monitoring\\_Resources](https://infrastructure.wiki.d4science.research-infrastructures.eu/production/index.php/Environmental_Monitoring_Resources)
- [9] D4Science FARM community resources [ONLINE]  
[https://infrastructure.wiki.d4science.research-infrastructures.eu/production/index.php/Fishery\\_and\\_Aquaculture\\_Resources](https://infrastructure.wiki.d4science.research-infrastructures.eu/production/index.php/Fishery_and_Aquaculture_Resources)