d4SCIENCE

Project acronym	D4Science
Project full title	**DI**stributed col**L**aboratories **I**nfrastructure on **G**rid **En**abled **T**echnology **4** **Science**
Project No	212488

Deliverable No DSA1.3a

Infrastructure Operation Report

July 2008

SEVENTH FRAMEWORK PROGRAMME
Research Infrastructures

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

DOCUMENT INFORMATION

Project	
Project acronym:	D4Science
Project full title:	**DI**stributed col**L**aboratories **I**nfrastructure on **G**rid **EN**abled **T**echnology **4 Science**
Project start:	1st 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:	DSA1.3a
Deliverable title:	Infrastructure Operation Report
Contractual Date of Delivery:	July 2008
Actual Date of Delivery:	20 August 2008
Editor(s):	Pedro Andrade
Author(s):	Pedro Andrade, Roberta Marque, Pasquale Pagano
Reviewer(s):	Paolo Fabriani
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:	1.1
Draft/Final	Final
Total number of pages: (including cover)	25
Keywords:	Production Infrastructure, Portal, VREs, Report

CHANGE LOG

Reason for change	Issue	Revision	Date
First version of the document	1	1.0	July 08
Integration of reviewer comments	1	1.1	July 08

CHANGE RECORD

Issue: 1 Revision: 1.0

Reason for change	Page(s)	Paragraph(s)
First version of the document	All	All

Issue: 1 Revision: 1.1

Reason for change	Page(s)	Paragraph(s)
Integration of reviewer comments	Several	Several

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

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
JRA	Joint Research Activity
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 on the current status of the D4Science production infrastructure that is being deployed by the project Service Activity. 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 the deliverable reports on the different nodes that compose the infrastructure by describing the status of these nodes and explaining the deployment, certification, monitoring, and support activities that lead to the availability of the infrastructure. Finally, this report presents the portal of the infrastructure giving details of the existing virtual research environments (VREs) and associated users communities.

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). This must be a well-supported infrastructure, running stable, tested and reliable software. The achievement of this production service has been planned in three major milestones:

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

This deliverable, being due on July 2008, has the main objective to report on the status of the production infrastructure after milestone MSA1.1. It describes the production infrastructure in terms of operations and usage.

From the operations perspective all partners have provided the resources that were defined in the beginning of the project. All sites are running gLite and gCube nodes as planned. Furthermore all partners are active in the different infrastructure organisational roles assigned to them.

Concerning the usage of the infrastructure, the portal has been deployed and configured. Different VREs have been created and can be made available to the Earth Monitoring community. An extra VRE for dissemination and training has also been prepared.

This deliverable presents this information in three main sections. Section 1 "Infrastructure Operation" presents the different types of nodes that compose the infrastructure and the status of the activities carried out by SA1 and SA2 work-packages to put these nodes in place. Section 2 "Infrastructure Usage" explains the virtual research environments (VREs) made available by the portal. It provides information about the VRE functionality, the data and tools associated to each VRE, and the communities and users allowed to access these VREs. Finally, section 3 "Conclusions" summarizes the deliverable by highlighting the most important achievements and difficulties reported in the deliverable.

1 INFRASTRUCTURE OPERATION

The D4Science production infrastructure is structured in three organisational layers:

- Management Centre (coordinator)
- Resource Centres (sites)
- Support Centres

The management centre is responsible for the management and the operation of the production infrastructure.

Resource centres are responsible for the deployment and operation of the different nodes of the production infrastructure. These nodes can be classified in three different categories:

- gLite nodes
- gCube nodes
- community nodes

Support centres are responsible for providing support to community end-users and operational support to site/portal managers on all issues related to the production infrastructure deployment and usage.

1.1 Overview

The D4Science production infrastructure currently counts with the participation of seven members of the consortium. These partners are involved in the different aspects of the infrastructure operation.

	Management Centre		Resource Centre		Support Centre	
	Planned	Actual	Planned	Actual	Planned	Actual
CERN	✓	✓			✓	✓
CNR			✓	✓	✓	
ESA			✓	✓		
NKUA			✓	✓	✓	
UNIBASEL			✓	✓		
ENG				✓		
USTRATH				✓		

Table 1 - Production infrastructure organisation

The status reported in Table 1 is inline with the plans defined in the beginning of the project in deliverables "DSA1.1a Procedures and Resources Plan" [1] and "DSA1.2a Middleware Deployment and Operation Support Procedures" [2]. Almost all partners are actively involved in the activities assigned to them. Furthermore, all resources planned for the end of June have been delivered by the different partners.

One deviation with respect to the original plan concerns the number of support centres involved. Until the end of June, only CERN acted as support centre. The reason for the non-participation of other partners as support centres was the limited number of support

tickets received. The support centre role is planned to rotate among different SA1 members to distribute the effort of managing such tickets. Having few tickets submitted, this rotation became unnecessary.

Concerning resource centres, the original plan foresaw the participation of four partners. However, with the decision to run two different releases of the gCube software for the milestone of end of June, two more sites (ENG and DBM-USTRATH) were requested to temporarily join the production infrastructure.

The following table gives an indication of the number of nodes dedicated to the infrastructure by the different resource centres.

	gCube Nodes	gLite Nodes
CNR	18	7
ESA	2	11
NKUA	22	6
UNIBASEL	6	4
ENG	2	1
USTRATH	1	0
TOTAL	**51**	**29**

Table 2 - gCube and gLite nodes contribution

Table 2 reports only on "gLite nodes" and "gCube nodes" since only these two types of nodes are described in this section. "Community nodes" are reported in section 2 since they relate to the data collections and tools exploited by the different user communities of the project.

Apart from nodes provided by the members of the consortium, the D4Science production infrastructure also foresees the possibility to exploit resources from external organizations or projects. A first example of this sort of collaboration has already been achieved with one site of the EGEE production infrastructure (Trinity College, Dublin) which gave permission to D4Science to use their gLite nodes.

1.2 gCube Nodes

gCube nodes are hosting nodes available to run gCube services. gCube is the software initially developed by the DILIGENT project now being extended in D4Science. The gCube software is composed by a special web service container, called gCube Hosting Node (gHN) and a set of WSRF specific services and libraries that provide the means to create, manage and exploit Virtual Research Environments (VREs).

1.2.1 Deployment

As explained in previous SA deliverables the deployment strategy of the gCube nodes can be summarized in the following steps:

1. Hosting nodes deployment
2. Core services deployment
3. Portal deployment
4. VRE services deployment

Steps 1 to 3 are under SA1 work-package responsibility since they provide the infrastructure core components that allow the deployment by SA2 of the gCube services

(step 4) implementing the management and operation of VREs. Currently several partners are contributing with gCube nodes as follows:

	Hosting Nodes	Core Services	Portal	VRE Services
CNR	✓	✓	✓	✓
ESA	✓			✓
NKUA	✓			✓
UNIBASEL	✓			✓
ENG	✓	✓		✓
USTRATH	✓			✓

Table 3 - gCube services contribution

To report on the gCube release currently deployed on these nodes it is first important to explain the evolution of the hosting node component. Starting from the existing DIILIGENT Hosting Node (DHN), developed during the DILIGENT project, a new gCube Hosting Node (gHN) has been developed in D4Science. The major improvement in the gHN is the integration of the new gCore framework. This framework packages the Globus Toolkit 4 container together with several libraries that facilitate the development of WSRF services. As a consequence, a refactorization of all existing gCube services was needed.

Due to the unavailability of the refactored version of some gCube VRE services, the Service Activity decided to revise the original plan. The adopted solution was to run two different releases of gCube:

- DHN-based release: this corresponds to the previous version of gCube (as released by the DILIGENT project). Even if not supporting the new gCore framework, this version of gCube provides the same level of functionality and is considerably more stable. It was already deployed and exploited in the DILIGENT pre-production infrastructure. This version of gCube can therefore still be used by the EM user community to perform the activities planned for MSA1.1.
- gHN-based release: this corresponds to the new version of gCube (as release by D4Science SA3). Even if few VRE services have been refactored by JRA and released by SA3, the gHN and gCube core services are already available. Despite the fact that no VREs can be created, the deployment of the new gCube Hosting Nodes and gCube core services was carried out to allow a quick and smooth transition to this new version of gCube.

This deployment plan was implemented and currently the different SA1 partners are providing both gHN-based and DHN-based gCube nodes. All partners are providing all the resources that were planned in the beginning of the project.

	DHNs	gHNs
CNR	15	3
ESA	0	2
NKUA	21	1
UNIBASEL	5	1

ENG	2	-
USTRATH	1	-
TOTAL	**44**	**7**

Table 4 - DHN/gHN distribution

The version of the gCube software deployed is:

- DHN-based nodes: DHN 1.0 (final DILIGENT release)
- gHN-based nodes: gHN 1.0.0 (first D4Science JRA release including gCore 0.3.1)

This current deployment scenario is foreseen to be in place during a short period of time while the re-factored gCube services are not released. As soon as they are officially released by SA3, their deployment will be scheduled in the production infrastructure and the DHN-based nodes will be gradually dismissed.

In order to support this transition phase, two new partners ENG and BDM-USTRATH, were requested to contribute with additional resources. They have been temporarily integrated in the operation of the production infrastructure until this transition period is over. Due to these circumstances these two partners are only providing DHN-based nodes.

It is important to report that this temporary deployment plan for gCube nodes, does not impact on the user communities exploitation plan since the level of functionality provided is the same. This new plan is inline with the original plan (as described in the project description of work).

Finally, with respect to the deployment of the portal, only the version from the previous gCube software is available. It has been deployed at CNR and is managed by a defined group of Portal Managers. Further details about the portal are provided in section 2.

1.2.2 Certification

The certification of a gCube node of the production infrastructure includes a number of tests to verify the correct functioning of the gHN. As a result of the certification, a gHN will be labelled as "Certified gHN", "Suspended gHN", or "Uncertified gHN". For further information about gCube certification details refer to deliverable "DSA1.2a Middleware Deployment and Operation Support Procedures" [2].

The certification procedure can only be applied to gHN-based nodes. This is due to the fact that only the gHN includes a monitoring component that periodically performs a local certification in the node. This component is not available in the DHN-based nodes.

Even if the D4Science certification procedure cannot be applied on DHN-based gCube nodes, these nodes were allowed to join the production infrastructure since they have been extensively tested and exploited before in the DILIGENT pre-production infrastructure.

Currently the 7 gHNs running in the production infrastructure are all "Certified gHN".

1.2.3 Monitoring

The monitoring of the gCube nodes is based on the gCube resources information published in the gCube Information System service. This information is visualized using the gCube monitoring portlet. As explained before, the infrastructure currently runs two different releases of gCube. Since the hosting nodes and services of these two releases are published in a different way, two instances of the portlet have been deployed:

- Monitoring portlet for DHN-based nodes
    - o http://portal.d4science.research-infrastructures.eu

  - the portlet is integrated in the official gCube infrastructure portal together with all other portlets provided by gCube
- Monitoring portlet for gHN-based nodes
  - http://dlib18.isti.cnr.it/monitoring
  - the portlet runs in a separate portal dedicated to this task

The current version of the monitoring portlet provides a flat view of the infrastructure. gCube Hosting Nodes, Services, Running Instances, and all other resources are seen from the same view. Future developments plan to have different views for the different resources of the infrastructure.

1.2.4 Support

The support to the operation of gCube nodes was implemented as described in deliverable "DSA1.2a Middleware Deployment and Operation Support Procedures" [2]. All support tickets related to the usage or operation of gCube nodes were managed using the TRAC[1] issue tracking system. In total 15 tickets related to gCube nodes were created.

	Infrastructure	Community
Open	3	0
Closed	10	2
Total	**13**	**2**

Table 5 - gCube support tickets

Some support tickets are still open. These tickets have been opened recently to address a number of problems related to the final configuration and delivery of the portal. These tickets are expected to be closed soon.

1.3 gLite Nodes

gLite nodes are computing and storage nodes running gLite software. gLite is the middleware provider 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.

1.3.1 Deployment

As introduced in section 1.1 there are five sites providing gLite nodes to the production infrastructure: CNR, ESA, NKUA, UNIBASEL, and ENG. The following table depicts which gLite services are provided by these sites (independently of their certification status).

	CE	WN	sBDII	SE	WMS	LFC	VOMS
CNR	✓	✓	✓	✓	✓		✓
ESA	✓	✓	✓	✓		✓	
NKUA	✓	✓	✓	✓	✓		

[1] https://issue.d4science.research-infrastructures.eu

UNIBASEL	✓	✓	✓	✓		
ENG						✓

Table 6 - gLite services distribution

All gLite services foresaw in the beginning of the project have been made available by the different SA1 partners involved. Furthermore, due the decision to run two different gCube releases in the infrastructure gCube nodes, as explained in section 1.2.1, there was the need to temporarily bring to the production infrastructure a second instance of VOMS. This new component was installed at ENG and the site temporarily added as production site.

From Table 6 it is also clear that the five sites provide all gLite services needed to create an independent gLite infrastructure. Nevertheless, these sites have joined the EGEE production infrastructure and extend it with more sites, providing computing and storage resources:

	Worker Node	Storage Element
CNR	4 CPUs	1.49 TB
ESA	14 CPUs	13.24 TB
NKUA	4 CPUs	0.27 TB
UNIBASEL	-	-
ENG	-	-
TOTAL	**20 CPUs**	**15 TB**

Table 7 - gLite nodes computing and storage resources

Almost all gLite sites of the D4Science production infrastructure run gLite version 3.1. There are some few exceptions where gLite 3.0 is still used. This situation however, does not bring any incompatibility problem.

With respect to the support for virtual organisations, all gLite nodes provided by D4Science are configured to support the following VOs:

- "d4science.research-infrastructures.eu": production VO of the D4Science project
- "ops": operations VO used by the EGEE operations teams

The "d4science.research-infrastructures.eu" VO has also been registered in the EGEE CIC[2] Portal and approved by the EGEE OAG[3]. This allowed to start negotiating with the EGEE VO Managers group the support by other EGEE sites of the D4Science VO. Some initial results have already been achieved and one site (Trinity College, Dublin) is now supporting the D4Science VO. This gives to the project the possibility to exploit plus 700 CPUs and 45 TB of disk space.

Finally, it must be reported the close collaboration that the five 4DScience gLite sites established with the different Regional Operations Centres (ROCs) of the EGEE project. The gLite deployment was extremely facilitated by the efficient support provided by the EGEE ROCs.

[2] EGEE Core Infrastructure Centre
[3] EGEE Operations Advisory Group

1.3.2 Certification

The certification of the D4Science gLite nodes follows the certification procedures established by EGEE for its production infrastructure. The execution of the certification tests is done by the EGEE ROCs. One gLite site is considered "Certified" when all its services are running properly and passed all certification tests.

Currently, 3 out of the 5 sites involved in the first milestone of the production infrastructure are "Certified". These sites are:

- CNR
- ESA
- NKUA

UNIBASEL has completed the installation of the all gLite services under its responsibility. However due to configuration problems in the gLite CE it is still being tested by the DE/CH EGEE ROC and cannot be classified as "Certified" yet.

As explained before, the gLite site from ENG is only temporarily part of the production infrastructure, so they did not undergo the certification procedure.

The status of the EGEE site certification can be monitored online via the EGEE GOCDB[4]:

- https://goc.gridops.org

1.3.3 Monitoring

There are several tools to monitor the EGEE production infrastructure nodes (and consequently the D4Science gLite nodes). Many of these tools share the same information source providing only different views over it. Such large number of tools covers many monitoring possibilities. The most relevant are:

SAM

- EGEE tool to monitor the availability of the gLite services and execute system tests
- https://lcg-sam.cern.ch:8443/sam/sam.py

GStat

- EGEE tool with online statistics of the infrastructure
- http://goc.grid.sinica.edu.tw/gstat

GridView

- EGEE tool to monitor the availability of the gLite services and query the infrastructure "Data Transfer" and "Job Status"
- http://gridview.cern.ch

GridMap

- EGEE tool providing a high-level graphical visualisation of the status of the infrastructure
- http://lxb2003.cern.ch/gm/gridmap.html

RTM

- EGEE tool for a 3D real time monitoring of the infrastructure
- http://gridportal.hep.ph.ic.ac.uk/rtm

All these different tools are used according to the different needs of the D4Science SA1 activities.

1.3.4 Support

The support to the operation of gLite nodes has been implemented as described in deliverable "DSA1.2a Middleware Deployment and Operation Support Procedures" [2]:

1. User support requests related to the usage of gLite were managed via the D4Science issue tracking system

[4] EGEE Grid Operations Centre Database

- https://issue.d4science.research-infrastructures.eu

2. Operations requests related to the deployment and upgrade of gLite were managed directly with the EGEE ROCs (either by email of via the EGEE global/regional helpdesks).

Concerning the support to the usage of gLite, only 3 support tickets were submitted and were monitored by SA1 support centres. The few support tickets received are related to the configuration of the D4science VO in some gLite services.

On the other hand, there was more activity in the support requests related to the deployment of gLite. During the first six months of the project, part of SA1 effort was dedicated to the initial deployment of gLite. This resulted in a number of support requests from all five sites to the different EGEE ROCs.

2 INFRASTRUCTURE USAGE

This section describes how the infrastructure presented in section 1 is being exploited by the D4Science members and in particular the project user communities. The usage of the infrastructure is explained by showing the portal of the infrastructure, the different types of infrastructure roles, the data sources accessible, and the specific VREs already created.

2.1 Portal

The D4Science portal provides access to end-user and administration functions in a unified way. It exploits the portal technology that is part of the gCube framework and it is based on GridSphere portal engine. The D4science portal is accessible at:

- http://portal.d4science.research-infrastructures.eu

The portal has been configured to allow two main types of users:

- Privileged users: can access the Users, VOs, and VREs management areas.
- Authorised users: can access existing VREs.

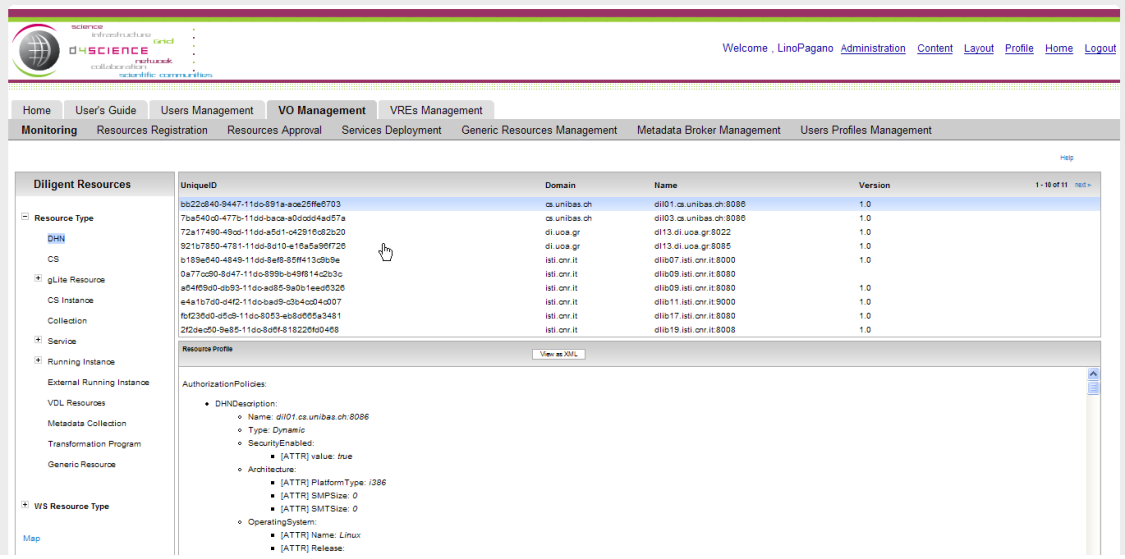

Figure 1 - D4Science portal privileged user area

Privileged users have one or more of the following roles:

Infrastructure Admin: this role grants the rights to invite, add, and remove users to the infrastructure; to assign and remove the VO Administrator role; to add and remove gHNs to the infrastructure; to monitor the status of gHNs of the infrastructure.

VO Administrator: this role grants the rights to invite, add, and remove users to the VO; to create and remove roles for the VO; to deploy and un-deploy services in the infrastructure; to register and remove gHNs and any other infrastructural resource; to define, register, modify, and remove metadata transformation and any other workflow used by the infrastructural services.

Resource Owner: this role grants the rights to register new resources in a VO. The resources are then approved or rejected by the VO administrator.

VRE Designer: this role grants the right to define a new VRE by using existing resources of the VO. The request are then approved or rejected by the VRE administrator.

VRE Administrator: this role grants the right to modify, approve, or reject the requests for the creation of new VREs.

Authorised users are users belonging to one or more VREs. A VRE provides a framework of applications, services, and data sources dynamically identified to support the underlying processes of research/collaboration/cooperation of a specific set of users belong to a community. Thus different VREs provide access to different set of functionalities operating over a different set of data sources. After the authentication of the user performed by means of a login operation the list of VREs that user is member of is presented. At any time a user can operate in a single VRE and, even if the switch among different VREs is done smoothly, each VRE presents a well defined space that does not interfere with other VREs. This prevents the contamination of a VRE with data and/or applications not selected by the VRE Designer and then approved by the VRE Administrator.

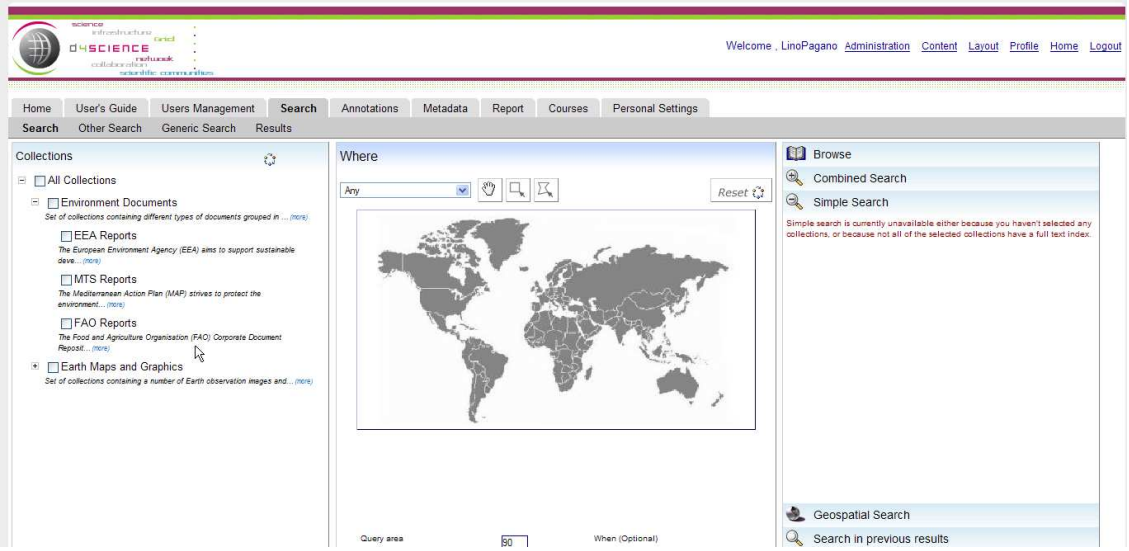

Figure 2 - D4Science portal authorised user area

2.2 Communities

The users and resources belonging to the Environmental Monitoring community represented in the D4Science consortium by ESA have been registered in one gCube Virtual Organization, called ImpECt.

The VO defines clearly and carefully what is shared, who is allowed to share, and the conditions under which sharing occur. More in details, this VO grants access to data resources provided by different international organizations and exploits non-gCube services that are maintained by third-parties.

In particular, it provides a uniform and integrated access to several community nodes. These nodes correspond to a number of data sources published by the following agencies and institutions:

EEA Reports: The European Environment Agency (EEA) aims to support sustainable development and to help achieve significant and measurable improvement in Europe's environment through the provision of timely, targeted, relevant and reliable information to policy making agents and the public.

MTS Reports: The Mediterranean Action Plan (MAP) strives to protect the environment and to foster development in the Mediterranean Basin. It was adopted in Barcelona, Spain in 1975 by 16 Mediterranean States and the EC, under the auspices of the United Nations Environment Programme (UNEP). Its legal framework comprises the Barcelona Convention adopted in 1976 and revised in 1995, and six Protocols covering specific aspects of environmental protection.

FAO Reports: The Food and Agriculture Organisation (FAO) provides access to the Corporate Document Repository that houses FAO documents and publications, as well as selected non-FAO publications, in electronic format.

Earth Images: The European Space Agency (ESA) offers a collection of earth images acquired from different instruments on-board of several Earth observation satellites.

Moreover the following data sources were collected by different sources of information, organized as collections, and registered to the ImpECt VO:

GeoNetwork Maps and Images: Geospatial information collected and harvested from a number of GeoNetwork nodes.

Chlorophyll Distribution Products: Data on Chlorophyll distribution and primary productivity for locating potential fish zones, sea surface velocities, suspended sediment movement, coastal landforms, coral reefs, etc. 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 is the base of the marine food chain and, therefore it is a good indicator of the abundance of life in a body of water.

Landsat7: Landsat satellites have taken specialized digital photographs of Earth's continents and surrounding coastal regions for over three decades, enabling people to study many aspects of our planet and to evaluate the dynamic changes caused by both natural processes and human practices. The purpose of the Landsat program is to provide the world's scientists and application engineers with a continuing stream of remote sensing data for monitoring and managing the Earth's resources. Landsat 7 is the latest NASA satellite in a series that has produced an uninterrupted multispectral record of the Earth's land surface since 1972. Applications include: urban change detection maps, land surface simulations and monitoring, deforestation mapping.

The ImpECt VO is therefore a collector of resources created and maintained by EEA, MAP, FAO, GeoNetwork, NASA, and ESA organisations by means of their own technologies and services. These resources are then homogenised and made interoperable through the gCube services in the D4Science infrastructure.

Moreover, in order to generate human consumable resources, the images and graphs belonging to the Chlorophyll Distribution Products and Landsat7 collections have been generated by running complex jobs submitted to the ESA infrastructural resources. The results of this execution have been then stored and further elaborated to make them interoperable with the other resources of the ImpECt VO.

At present, the resulting ImpECt VO provides seamless access to heterogeneous resources that can be exploited to create courses and reports. However, all the data sources are not automatically synchronised with the external data sources but it is up to the VO Administrator to perform this synchronisation activity periodically. This activity is time consuming and prone to error. For this reason a new set of functionalities have been required to the Joint Research Activities team in order to transfer this human-based process to an infrastructural service that can perform it in a complete autonomic mode. This requirement has been accepted and the new service implementing these functionalities is in advanced implementation stage.

2.3 VREs

The ImpECt VO provides access to three different VREs created by exploiting the capabilities of the VRE Management area of the portal:

ImpECt Digital Library: this VRE provides access to all data sources available in the ImpECt VO. It offers full text, geospatial, structured, and semi-structured search and browse capabilities over the heterogeneous data sources. It permits the annotations of objects, the creation of courses, the design and generation of reports. It allows saving temporary results, objects, and metadata in a personal area called "basket" that it is maintained across different sessions. The goal of this VRE is to provide access to a broad and heterogeneous set of users that range from expert researchers capable to elaborate data and create new products to data consumers that want to access data without loosing time in understanding different and not uniform interfaces, protocols, access rights, policies and procedures.

ImpECt Vegetation: this VRE provides access to the same set of functionalities of the ImpECt Digital Library over a subset of the available data sources and courses. This VRE is specialised to serve researchers in the area of Vegetation management. The products generated by this VRE become automatically part of the ImpECt Digital Library VRE.

ImpECt Demo: this VRE is used for dissemination and training. It provides access to the same resources available in the ImpECt Digital Library VRE to a larger community of users with fewer restrictions for their participation. The users of the ImpECt Demo VRE cannot store annotations, persist courses or reports. Their personal area (their basket) is not maintained across sessions since any ImpECt Demo session runs in a sandbox that it is destroyed at user logout.

The three VREs can use 66 gCube services and 34 software components provided by third-parties. At time of writing this report, they exploit 130 Running service Instances (RIs). However this number can change along time to better satisfy the needs of the VREs. The number of RIs grouped per service classes is reported in Table 8.

Service Class	# RIs	Site
Virtual Organization Management	15	CNR, NKUA, UNIBASEL
VRE Management	18	CNR, NKUA, UNIBASEL
Information System Management	3	CNR
Content Management	4	UNIBASEL
Content Source Description and Selection	2	NKUA
Metadata Management	4	CNR, NKUA
Data Fusion	1	NKUA
Annotation	1	CNR
Search	37	CNR, NKUA
Feature Extraction	0	-
Index Management	25	CNR, NKUA
Personalization	2	NKUA
Process Management	15	CNR, NKUA
Portal	3	NKUA
Total	**130**	

Table 8 - RIs distribution in the ImpECt VO

3 CONCLUSIONS

The D4Science production infrastructure has been deployed and can be exploited by the Earth Monitoring user community and the project dissemination and training teams. MSA1.1 can therefore be considered as accomplished.

To reach this point a series of difficulties had however to be overcome (some are still present) and are now reported from the perspective of future improvements:

- The new deployment plan for gCube, bringing together DHN-based and gHN-based nodes, added additional effort in the operation of the production infrastructure. Even if no new procedures had to be defined, the management of two different software releases is always a costly activity from the management and operational perspective.
- The certification of the gLite nodes in UNIBASEL site was not completed. The site is still in close communication with the EGEE ROC to overcome this situation.
- The gCube and gLite nodes in NKUA site were provided with some weeks delay. This situation was caused by late delivery of new hardware. Temporary installations were done on virtual machines and a gradual transition to the new hardware is now being performed.
- A number of unforeseen secondary problems have been encountered in the set-up, configuration and operation of the D4Science portal. In particular problems related to the management of the credentials, reallocation of nodes, organisation of the virtual organization users and groups, set up of a demonstration VRE, etc. Adequate actions were taken and appropriate support was provided by the development teams. Nevertheless the process was quite time-demanding and prevented the official release of the portal in production by the end of June, adding about 3 weeks of delay.

From another perspective, not only the achievement of the milestone must be reported but also a number of other successful facts:

- Despite the changes introduced in the deployment plan of gCube nodes, there was no impact on the project user communities that were provided with the functionality planned in the beginning of the project.
- Even if not planned in the beginning of the project, two more partners are temporarily providing resources to the production infrastructure (both gCube and gLite nodes).
- One EGEE gLite site adopted the D4Science VO and provided to the project access to a considerable extra number of resources.
- Although not originally planned, a new VRE for dissemination and training was created. This VRE corresponds to the needs expressed by the project dissemination and training activities that requested access to the production infrastructure for their activities.
- Setting up a production infrastructure is a complex task. In the D4Science case, this experience was definitely challenging but at the same time extremely fruitful. It allowed to better understand the weaknesses of the infrastructure and learn how to inprove the future operation of the D4Science production infrastructure.

REFERENCES

[1] D4Science DSA1.1a Procedures and Resources Plan
http://bscw.research-infrastructures.eu/bscw/bscw.cgi/d79694/D4Science_DSA1.1a_M1.pdf

[2] D4Science DSA1.2a Middleware deployment and operation support procedures
http://bscw.research-infrastructures.eu/bscw/bscw.cgi/d85323/D4Science_DSA1.2a_M4.pdf