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 D2.1a

Quarterly Report

April 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:	D2.1a
Deliverable title:	Quarterly Report
Contractual Date of Delivery:	April 2008
Actual Date of Delivery:	19 May 2008
Editor(s):	Donatella Castelli (CNR)
Author(s):	Pedro Andrade (CERN), Donatella Castelli (CNR), George Kakaletris (NKUA), Johannes Keizer (FAO), Pasquale Pagano (CNR)
Reviewer(s):	N/a
Participant(s):	CNR, NKUA, CERN, FAO
Work package no.:	NA2
Work package title:	Scientific and Technological Coordination
Work package leader:	CNR
Work package participants:	CNR, NKUA, CERN, ESA, FAO
Est. Person-months:	1
Distribution:	Confidential
Nature:	Report
Version/Revision:	1.0
Draft/Final	Final
Total number of pages: (including cover)	25
Keywords:	D4Science Activities; D4Science Events

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

TABLE OF CONTENTS

LIST OF TABLES

SUMMARY

This report describes the achievements of the D4Science activities with respect to the planned objectives for the period January - March 2008. Section 1 provides a summary view of these achievements grouped into four activity areas: *coordination and management*, *outreach*, *service* and *joint research*. Section 2 details these achievements by reporting on the progresses done in the context of each activity area per constituent work package. Section 2.5 presents performance indicators measuring various aspects of the performed activities. Section 4 concludes this periodic report by summarising the main events occurred in the reporting period.

1 SUMMARY OVERVIEW OF THE ACHIEVEMENTS OF THE PERIOD

1.1 Coordination and Management Activities

In this reporting period the Coordination and Management Activities have reached the following achievements:

- Collection of A Forms and signature pages to the Consortium Agreement for all partners except for FAO and ESA;
- Launch of project via a week of meeting at ERCIM in Sophia Antipolis, 7-11 January;
- Formation of the PMB (MNA2.1);
- Setting up of tools for supporting intra-project cooperation;
- Distribution of the first advance payment, to all partners except FAO and ESA;
- Definition of the project's Quality Assurance procedures, including the establishment of a wiki for procedures, indicators, documentation and activity monitoring;
- Validation and approval of the deliverables due in this period;
- Detailed Activity Plan (MNA2.2);
- Organization and management of the first two plenary Technical Committee meetings;
- First meeting of the External Advisory Board held on 26 March.

1.2 Outreach Activities

In this reporting period the Outreach Activities have reached the following achievements:

- The D4Science website MNA3.1a (http://www.D4Science.eu) was released;
- The D4Science project was presented at several events;
- Participation agreement reached between FAO and three important external user community partners: NAFO, ICES and NEAFC for the GW4SA VRE;
- Participation agreement for the FCPPS VRE reached with the two main internal producers of country-level data, FIMA and FIP together with strong support from the FIP director;
- Requirements collaboration between FAO and World Fish Center;
- Invitations sent to 40+ proposed external user Fishery community partners;
- Contribution to the brochure for the area of Scientific Data Infrastructures;
- Several internal requirements gathering meetings and presentations held with FAO staff.
- Cross-community meeting between FAO and ESA with the participation of CNR and ENG representatives held at FAO in March 2008;
- Validation and assessment of the web portal released by the DILIGENT project. Analysis of DILIGENT achievements reached with the implementation of the earth science scenario;
- Scientific research topics for EM VREs definition have been targeted;
- Identification of potential EM data sources according to targeted scientific research topics;
- Investigation activity on current web portals and digital libraries implementations providing access to Earth science information;
- Agreement reached between CNR, NKUA and ESA on the proposed plan for system adoption by EM community;
- Positive feedback on the project collected from ESA staff participating at the ESA/ESRIN event "3rd Grid & e-Collaboration workshop for the Earth Science Community".

1.3 Service Activities

In this reporting period the Service Activities have reached the following achievements:

- Plans, procedures, and deliverables have been produced for the integration, testing, and distribution of gCube/gCore, and the deployment of the production infrastructure;
- A set of collaboration tools were selected and configured to support the SA activities;
- Initial contacts with the project user communities were established and a concrete work plan was defined.
- The release cycle to build, package, test, and distribute the gCube and gCore software was established. Builds over development code started to be executed;
- The project production infrastructure was been planned. Its structural organisation, procedures, and tools were clearly indentified. A concrete deployment plan until June 2008 was established. The deployment and certification of gLite nodes started;
- Quality metrics to monitor the work of the Service Activity were indentified.

1.4 Joint Research Activities

In this reporting period the Joint Research Activities have reached the following achievements in order of their realisation:

- Requirements analysis and system design of gCore (3rd iteration);
- Advanced release of Index Service;
- Compilation of detailed dependency graph;
- Analysis of preliminary user requirements and compilation of system-oriented requirements;
- Delivery of tools for JRA activities;
- Overall planning of JRA activities;
- Design of Core Services;
- Design of Information Organisation Services;
- Design of Information Retrieval Services;
- Design of Presentation Services;
- Implementation and first internal release of gCore;
- Re-factoring of several Core Services for gCore adoption;
- Re-factoring of some Information Retrieval Services for gCore adoption;
- Issue-fixing in Information Retrieval Services and Components.

2 ACTIVITIES PROGRESS OF THE PERIOD

2.1 Coordination and Management Activities Progresses

Activities performed

This first reporting period of the project has been characterised by "standard" administrative and financial activities. In particular, the organization of the kick-off meeting (Technical Committee, Members General Assembly and Quality Assurance Task Force) at ERCIM, Sophia Antipolis, France from 7-11 January 2008 has represented the official start of the project. However, before the official start of the project actions needed to collect the duly signed A Forms to the EC Grant Agreement which entered in force on 14/12/2007, and the signature pages to the D4Science Consortium Agreement (Version 1 November 2007) have also been performed.

With respect to the administrative activities, reception and transfer of pre-financing to consortium members has been accomplished on 14 February; 35% of total financing have been transferred to all beneficiaries except for FAO and ESA.

With respect to the consortium operation, the following set of actions has been performed. The Project Management Board and the Quality Assurance Task Force (QATF) have been formed in conjunction with the project kick-off. In particular, a first QATF meeting has been held the day after the kick-off. During this meeting the methodology for defining the quality parameters and their measures has been identified.

The set of tools needed to support intra-project cooperation, e.g. mailing lists, shared workspaces and project calendar, have been deployed and tuned to satisfy the project needs.

With respect to the Quality Assurance, the QATF have been active since the early phases of the project. This task force leaded the activities conducting to the identification of the set of indicators capturing the advances of the rest of the project activities, and the tools and procedure needed to gather them. Moreover, the QATF in strict cooperation with the Project Director drafted the structure and the content of this reporting deliverable as well as the procedure governing the D4Science deliverable production. All these guidelines have been described in a wiki based web service and the relative procedures have been implemented by exploiting the available tools, e.g. the procedure governing the deliverable production have been implemented in terms of workflows supported by the shared workspace service (BSCW). Thanks to the support of the deployed tools the upkeep of the above wiki based web service become feasible, the QATF members have been continuously notified of project advances and these advances have been made public through the service. In particular, the project has defined the procedure for the preparation, review and acceptance of deliverables. This procedure has already been put in place for the deliverables due in this period.

The first External Advisory Board (EAB) meeting was held on 26 March in Tirrenia (Pisa), Italy. Due to the unavailability of Dr. Frank Rijsberman to participate in this meeting the three directors of the project decided to invite a fourth EAB member. The nominated member is Dr. Edward Vanden Berghe, current Executive Director of the Ocean Biogeographic Information System (OBIS).

A document that describes the project objectives and its planned activities was prepared and distributed to the EAB members before the meeting.

The project has been monitored from the scientific and technical point of you. A detailed activity plan has been elaborated and put in action since the early phases of the project. The activities performed by the Technical Director (TD) in conjunction with the Project

Director (PD) represented the glue among the different activities. The adoption of common tools, compatible procedures, and shared strategies in the horizon of the project milestones and duties were the first aim and achievement of the implemented activity plan. The clear identification of the areas of interaction across activities and the actions needed to smooth misalignments among them have been the focus of the daily work performed by the TD. This activity, simplified by the monitoring tools put in action in this first period of the project, has benefited from the active collaboration of the Activity Managers. It has been materialised by daily exchange of information and organisation and/or participation to phone conferences and meetings. In particular, the 2008 scheduling of the Integration, Coordination, and Planning meeting, also known as Technical Committee (Tcom) meeting, was elaborated by the TD and communicated at the first Tcom event organised in conjunction with the project kick-off held in Sophia Antipolis in January 2008. This schedule represents the main steps along which the project will implement its plans and corrective actions. Two Tcom meetings were held in the first three months: 8-10 January (Sophia Antipolis, France) and 17-19 March (Pisa, Italy); three will follow according to the following schedule: 26-31 May (Thira, Greece); 8-13 September (Basel, Switzerland); 24-28 November (location to be identified).

Problems Encountered/Deviations from Plan and Corrective actions taken

- At the time of the signature of the Grant Agreement by the Coordinator and the European Commission (14 December 2007), it was assumed that the special clause no. 3, formulated for the United Nations organizations, would be accepted by FAO. However, the Grant Agreement did not provide for the following: the governance by general principles of law to the exclusion of any single national system of law. It was after the entry in force of the Grant Agreement that ERCIM was informed by FAO of the inability to accede to the Grant Agreement due to the content of the special clause no. 3, and discussions began with the Commission highlighting the need for continued negotiation between FAO and the European Communities with regard to the special clause(s) provided for the United Nations organizations.
- The Consortium Agreement was written to meet the special requirements of FAO, yet ESA would not accept the conditions stated in the Consortium Agreement because they are not in line with the Grant Agreement. Therefore, ESA declined to sign the Consortium Agreement, nor the Grant Agreement until a resolution is achieved with FAO.
- The general conditions of the Grant Agreement were released after the finalization of the D4Science Description of Work. Unlike in FP6, deliverables are due on the last day of the month of the "due date". Thus, the DoW stated delivery dates for deliverables that did not take into consideration the time required for performing a thorough internal review. As a result, the project work plan has been advanced by 6 weeks in order to produce deliverables on time according to the Technical Annex. The only exception is the delivery of the quarterly progress reports, as their production can not begin any earlier than after the closure of the period to be documented.

Achievements

- Collection of A Forms and signature pages to the Consortium Agreement for all partners except for FAO and ESA;
- Launch of project via a week of meeting at ERCIM in Sophia Antipolis, 7-11 January;
- Distribution of the first advance payment, to all partners except FAO and ESA;
- Formation of the PMB (MNA2.1);
- Setting up of tools for supporting intra-project cooperation;
- Definition of the project's Quality Assurance procedures, including the establishment of a wiki for procedures, indicators, documentation and activity monitoring;
- Validation and approval of the deliverables due in this period;
- Detailed Activity Plan (MNA2.2);
- Organization and management of the first two plenary Technical Committee meetings;
- First meeting of the External Advisory Board held on 26 March.

2.2 Outreach Activities Progresses

Activities performed

The Outreach activities was mainly dedicated to the definition of the communication strategy, aimed to vehicle the project's key messages, and to the aggregation of the communities and their data and tools, aimed to build consensus around the project's advances and novelties. The activities performed in this initial stage of the project can be summarised as follow:

- The project logo has been designed (available M1) and distributed in different formats and sizes through the BSCW, i.e. the project shared workspace;
- The project web site has been developed based on Drupal (Content Management System). It includes background information about the project, lists news and events, links to partners' web sites and includes search functionalities. Tests have started on the integration of Drupal and BSCW to display public Calendar and documentation;
- D4Science Flyer has been developed/printed/distributed at various occasions;
- An official press release has been prepared by ERCIM with ESA and FAO's inputs. It has not been published yet;
- A poster "D4Science: E-Infrastructures for Fisheries Resources Management." for the International Conference On Marine Data and Information Systems IMDIS in April 2008 has been prepared by NKUA with NA5's contribution;
- DNA3.1a Communication and Dissemination Plan has been prepared following the deliverable workflow. The review activity is ongoing;
- PowerPoint and Word templates for deliverables have been developed and distributed through the BSCW, i.e. the project shared workspace.
- D4Science has been added in the EGI Knowledge Base under the main European e-Infrastructure projects (http://knowledge.eu-egi.eu/index.php/Projects);
- D4Science was presented at several external events:
  - the 3rd Grid & e-Collaboration workshop for the Earth Science Community, ESA/ESRIN, January 2008;
  - First DRIVER Summit, January 2008;
  - 3rd EGEE User Forum, February 2008.
- Requirements collaboration between FAO and World Fish Center;
- Invitations sent to 40+ proposed external user community partners;
- Several internal requirements gathering meetings and presentations held with FAO staff;
- Cross-community meeting between FAO and ESA with the participation of CNR and ENG held at FAO in March 2008. PowerPoint presentations made by FAO/ESA/CNR;
- Internal project presentations and informal discussions on common objectives held with ESA staff;
- Further development of the Aquamaps website and packaging design for integration in the platform;
- Continuous production of maps of Aquamaps with improvement of algorithm for narrow-wide distributed species;
- Development of a cache that merges FishBase and SeaLifeBase data (marine fishes and non-fish invertebrates);
- Exploration to produce new services around these maps: biodiversity lists, integration of Global Climate Change scenarios, cross-ocean and coastal transect diversity;
- First contacts with OBIS, BFAR (Philippines), SEAFDEC (ASEAN);
- Development of a proposal jointly with OBIS to speed up the process of map producing;
- Validation and assessment of the web portal released by the DILIGENT project. Analysis of DILIGENT achievements reached with the implementation of the earth science scenario;

- Identification of detailed scientific research topics to include in project's VREs.
- Investigation activity on current web portals and digital libraries implementations providing access to Earth science information;
- Internal discussion on the proposed plan for system adoption by EM community held between CNR, NKUA and ESA.

Problems Encountered/Deviations from Plan and Corrective actions taken

- M2 "report" deliverable DNA3.1a has been delayed partly due to the late adoption of the 'Review process". The review of the document is still ongoing.
- M3 "other" deliverable DNA5.1 is delayed due to administrative problems. The content of the deliverable and the tool to be used to manifest it have been agreed with the appointed project reviewer. It has been rescheduled for M4.

Achievements

- The D4Science website MNA3.1a (http://www.D4Science.eu) was released at the beginning of M2;
- The D4Science project was presented at several events;
- Participation agreement reached between FAO and three important external user community partners: NAFO, ICES and NEAFC for the GW4SA VRE;
- Participation agreement for the FCPPS VRE reached with the two main internal producers of country-level data, FIMA and FIP together with strong support from the FIP director;
- Requirements collaboration between FAO and World Fish Center;
- Invitations sent to 40+ proposed external user Fishery community partners;
- Contribution to the brochure for the area of Scientific Data Infrastructures;
- Several internal requirements gathering meetings and presentations held with FAO staff.
- Cross-community meeting between FAO and ESA with the participation of CNR and ENG representatives held at FAO in March 2008;
- Validation and assessment of the web portal released by the DILIGENT project. Analysis of DILIGENT achievements reached with the implementation of the earth science scenario;
- Draft vision and scope documents produced by FAO for the FCPPS and GW4SA VREs;
- Scientific research topics for EM VREs definition have been targeted;
- Identification of potential EM data sources according to targeted scientific research topics;
- Investigation activity on current web portals and digital libraries implementations providing access to Earth science information;
- Agreement reached between CNR, NKUA and ESA on the proposed plan for system adoption by EM community.
- Positive feedback on the project collected from ESA staff participating at the ESA/ESRIN event "3rd Grid & e-Collaboration workshop for the Earth Science Community".

2.3 Service Activities Progresses

Activities performed

During these first three month of the project most of the work performed by the Service Activity was on planning the activity. In particular:

- The 3 deliverables expected for this period (DSA1.1a, DSA3.1, and DSA3.2) were successfully produced. These deliverables define: (1) the procedures and resources allocated to the production infrastructure, (2) the software release cycle, and (3) the testing activity. Two deliverables (DSA1.1a ad DSA3.2) were approved by the PMB and sent to the European Commission while DSA3.1 is under final approval.
- Different procedures related to the management of the software lifecycle and to the operation of the production infrastructure were defined. These procedures were

discussed by email, phone conferences, and at the 1st and 2nd D4Science TCom meetings. Most of them are described in the deliverables above.

To support the above procedures and facilitate the communication between the SA work-packages and between the SA and other project Activities, several collaboration tools were adopted or implemented. Most of the tools were selected at project level to allow an easier and more efficient communication between all project Activities:

- BSCW: documentation repository for all SA related documentation (including deliverables and deliverables review workflow).
  http://bscw.research-infrastructures.eu/bscw/bscw.cgi/54982
- TRAC: issue tracking system to manage infrastructure support request, integration and testing bugs, certification bugs, and release cycle tasks.
  https://issue.d4science.research-infrastructures.eu
- ETICS: distributed build and test system to build, package, distribute, and apply deployment and functional tests on the gCube and gCore software.
  https://etics.cern.ch:8443/eticsPortal
- Wiki: collaborative website hosting the public pages of the "Infrastructure Operation" and "Integration, Testing, and Distribution" work-packages.
  https://infrastructure.wiki.d4science.research-infrastructures.eu
  https://integration.wiki.d4science.research-infrastructures.eu

In addition, SA2 performed an analysis of the tools adopted by the project for the management of the communication and exchange of information within the work-package. Among them, BSCW and TRAC offer features to manage discussion forum. The former allows to create and to use one or more blogs; the latter allows to exploit an 'issue' request management feature. Even if the two approaches appeared equivalent being capable to be adopted to fulfil the needs of the project, TRAC was selected since it adds to the standard capabilities a set of useful characteristics. Among the others, it provides a mechanism to customize its behaviour, allows to store structured information, provides a rich set of search facilities, and includes a native mechanism to prioritize and organize the user requests. Moreover, it does not appear more complex than BSCW. SA2 configured TRAC properly and defined the process to be adopted any time a request is opened by the user communities.

Regarding ETICS, SA3 requested the creation of two new project entries (gCore and gCube) in the ETICS System. The initially empty projects were populated with first-level structure (subsystems) according to the architecture inherited from the DILIGENT project.

It's important to note that SA3 deployed and is maintaining not only the SA wikis, but all wiki instances available for the different areas of the project: SA1 Infrastructure, SA3 Integration and Testing, NA1 Quality Assurance, JRA gCube and gCore Documentation.

Another important activity performed during this reporting period, were the initial discussions with the project user communities. In particular, SA2 opened a discussion with both user communities by participating in all phone conferences and meetings. Being responsible to 'translate the user requirement into a plan', SA2 explained the type of requirements expected, the desired level of details, and the needs to have preferably atomic requests.

In addition, SA2 explained to the user communities how to use TRAC. This activity was performed with a number of communications, all intended to clarify doubts and remove potential obstacles in its adoption.

The initial discussions with the user communities allowed a better understanding of some of their initial requirements. For these requirements, a plan of activities is by now ready to be implemented and it will be used as soon the requirement will be formally expressed by the users.

During this first quarter, a number of tasks related to the integration, testing, and

distribution of the software developed by the Joint Research Activity (JRA) were also carried out.

In particular, SA3 held some internal meetings to define the project Release Cycle, clearly identifying the actors involved in this process, the interaction among them and the collaboration tools to be exploited. The plan was presented to the involved partners at the 2nd TCom meeting held in Pisa. This information has been collected in the project deliverable DSA3.1.

As planned, during the month of March, SA3 deployed and configured a build server to run project-level builds using the ETICS system. The server is currently running nightly builds of both gCore and gCube systems and makes build reports available to D4Science developers through a web-based reporting application.

To prepare the deployment testing activity, and understand the potential of new ETICS functionalities, an experimentation using the new ETICS co-scheduling remote testing functionality was been conducted.

The SA3 testing team defined guidelines, software usage, responsibilities, and thoughts on how different kinds of testing that can be executed and performed over gCube and gCore systems. This information has been collected in the project deliverable DSA3.2.

For each gCube and gCore subsystem one manager has been identified. The Subsystem Manager is responsible for the communication with SA3 and to monitor and be informed about bugs and activities of the components belonging to his/her subsystem.

More details about the integration, testing, and distribution activities can be found in the "Integration and Testing" wiki page:

- https://integration.wiki.d4science.research-infrastructures.eu

With respect to software documentation, the material produced by the DILIGENT project was moved to two new wiki sites provided by SA3. These sites are the starting point for the production of the gCube and gCore user, administrator and developer guides. The wiki sites are available at:

- https://technical.wiki.d4science.research-infrastructures.eu
- https://wiki.gcore.research-infrastructures.eu

Progresses were also visible in the setting up of the production infrastructure. In the initial period of this quarter, most of the effort was dedicated to the definition of the organisation of the infrastructure and the planning of the resources allocated to it.

The D4Science infrastructure was organized in three types of centres: Management Centre (MC), Support Centre (SC), and Resource Centre (RC). The role of SC and RC was assigned to all SA1 partners while the MC role was assigned to CERN. Each Resource Centre can host different type of nodes: gCube nodes, gLite nodes, and Community nodes. More details can be found in the project deliverable DSA1.1a.

The deliverable also describes how the SA1 operations activities were organised in five areas of work: installation and upgrade, site certification, user and operational support, site security, and monitoring. Different procedures and tools were established for each area.

With respect to resource allocation, concrete plans for MSA1.1 have been established. The production infrastructure on M6 will be composed by 4 Resource Centres (CNR, ESA, NKUA, and UNIBASEL) which will provide 54 CPUs and 12,800 GB of storage. These RC will provide gCube nodes running gCube services and gLite nodes which will join the EGEE production infrastructure. A deployment plan for the resources allocated to the infrastructure was discussed through a number of phone conferences and presented to the project during the 2nd TCom meeting.

For each Resource Centre (aka site) one manager has been identified. The Site Manager is therefore responsible for the installation, maintenance, and support of the gLite, gCube, and community nodes that compose his/her site.

The installation of gLite nodes started in the month of March, and some of the sites are

already being certified to be accepted by the EGEE infrastructure. At the same time, discussions with EGEE started about the registration of the "d4science.research-infrastructures.eu" VO and the possible support from other EGEE sites to this VO.

More details about the production infrastructure can be found in the "Infrastructure" wiki page:

- https://infrastructure.wiki.d4science.research-infrastructures.eu

Finally, it should also be reported the work done by the Service Manager and SA work-package Leaders in the identification of quality metrics related to the SA activity. These metrics were sent to the project Quality Assurance Tasks Force (operating in NA1 activities). The SA indicators were grouped in four main areas:

- Infrastructure Size
- Infrastructure Usage
- Infrastructure Support
- Software Release

These indicators will be collected during the lifetime of the project and will be included in the project quarterly reports (cf. Section 2.5).

Problems Encountered/Deviations from Plan and Corrective actions taken

None

Achievements

- Plans, procedures, and deliverables have been produced for the integration, testing, and distribution of gCube/gCore, and the deployment of the production infrastructure;
- A set of collaboration tools were selected and configured to support the SA activities;
- Initial contacts with the project user communities were established and a concrete work plan was defined.
- The release cycle to build, package, test, and distribute the gCube and gCore software was established. Builds over development code started to be executed;
- The project production infrastructure was been planned. Its structural organisation, procedures, and tools were clearly indentified. A concrete deployment plan until June 2008 was established. The deployment and certification of gLite nodes started;
- Quality metrics to monitor the work of the Service Activity were indentified.

2.4 Joint Research Activities Progresses

Activities performed

The activities performed during the 1st 3 months of the projects fell in the following major 4 classes:

1. Preparation of JRA activities
2. Establishment of fundamental development entities
3. Differential design of services and components
4. Re-factoring/fixing of existing code

These are presented in detail below.

Preparation of JRA activities

Preparation of JRA activities started with the thorough analysis of the gCube code, which was delivered at the end of DILIGENT project, in the direction of meeting the objectives of D4Science. The evolution during the last 7 months of DILIGENT required that this analysis was redone, especially after further knowledge on preliminary D4Science user requirements was obtained. As a result a set of system-oriented requirements was assembled and validated by the JRA members and propagated back to the users for pre-

validation.

This initial stage of preparation was done in tight collaboration with other activities, such as Service Activities for managing the coupling of tasks at the boundary crossing points of releases, and Networking Activities for obtaining requirements and delivering information for the anticipated progress of JRA work to Training.

Having in hand the preliminary set of requirements, and an overall time plan for the activities, a detailed dependency graph was sketched for handling the complex implementation interdependencies. As such, it is a tool for validating the feasibility and the limits of implementation against the hard time-constraints of the project Description of Work. The final detailed plan, which guides not only the first release but also subsequent cycles of implementation, has been announced and verified. Explicit dates of delivery are to be released shortly, in order to take into account the experience of the re-factoring activities exercised during these first three months: the introduction of gCore was foreseen and a long-term return of investment is guaranteed, yet the impact at the initial stage of the project could not be calculated a-priori precisely, thus scheduling of the plan in terms of time could not come in earlier.

A number of tools were set-up for supporting the JRA tasks, in tight collaboration with Service Activities (and secondarily the Networking Activities) in order for a common environment to be built. TRAC and Subversion (SVN) are the main new instruments at the disposal of developers for everyday work, while the pre-existing tools were brought to the new context (ETICS, BSCW and traditional tools, such as mailing lists and WiKis).

The tools were gradually populated with content starting with "wishes and promises" which, when validated, evolve into requirements / specifications and currently they in the stage of collecting low level requirements which occur from component level design and implementation.

Last but not least, a number of procedures and metrics have been elaborated for measuring the quality of JRA and several are already adopted, while others are expected to be adopted in the upcoming period. The review of components and the high-level metrics of system performance are two examples.

Establishment of fundamental development entities

gCore, the new layer that removes significant part of repetition of code among gCube software components and simplifies the development of new ones, has undergone pre-design stages before the beginning of D4Science, led by CNR and USG (all other's contributing to design and concept's development). Adoption of gCore along with targeted design and re-factoring are the main vehicles for achieving the required stability and performance that will drive a production-level infrastructure. In this initial period of D4Science, bringing gCore to life was one of the two early-implementation objectives. After designing its final aspects, early in the project, the first release of gCore was fully factored and released in time to the rest of the development teams for use in their components.

As part of gCore, a large number of Core Services have been re-factored in order to support it effectively for its initial release. This includes services enabling the infrastructure and apart from implementation changes, the break down in components has been modified to better match the needs of the new approach.

Driven by the informally communicated user requirements and a large analysis of the evolving technology in the domain of user interfaces, the need to build a layer that will facilitate the implementation of user applications, lifting off the burden of implementing the various standards and conventions of gCube, has been identified. Similar terms of re-factorisation to the user-side components (portlets, servlets, etc) are anticipated and the gains will be of the same overall direction. An additional gain at this point will be the decoupling from the technology in the boundary of system logic - user interface.

The derived Application Support Layer has been designed to be a product of large re-factoring of existing user interface components. Its design is already through and is expected to be released also quite ahead of the first software release, although no user

interface changes are planned for the first deployment of the system, according to the current plan.

Differential Design of services and components

After the analysis of the initial informally communicated user requirements, the in-depth analysis of the current state-of-the-art and associated trends, and the evaluation of gCube status a stage of differential Design for the whole gCube platform was inserted to the plan.

Desktop analysis showed new opportunities for meeting user requirements (such as OAI-ORE), handling issues of the system (JSR286 and new portal engines) and bringing forth new features and capacities (e.g. versioned package repository).

As a result of this short design stage, which was based on DILIGENT architectural artifacts, new goals and approaches were conceived, some of which are by large guided by user requirements, and their application-specific implementations to be performed as part of JRA2 and JRA3. Among the most important goals we find the following ones:

- Live Complex Documents, which capture several aspects of the concept of document: from designing reports to persisting instances of documents. The new, currently labeled, "Virtual" Object is product of the integration of Complex Live Documents and Data Source Abstraction.
- State-of-the-Art approaches for interfacing with content containers with several different degrees of integration, ranging from full migration to mere links.
- Data Source Abstraction Layer that leverages access to information containers, lifting barriers for information and data handling in the domains of access, retrieval, definition and persistence.
- Application Support Layer, that allows user-centric applications to be built on top of an Infrastructure-centric system like gCube and solves cross-platform-wise the issues of session management and interfacing.
- Infrastructure-based support for versioning not at the level of package, but rather at the level of invocation, an un-preceded innovation in the SOA.

The handling of the above goals leads to a long term implementation which spans significant parts of the duration of the project. As such they could impact the progress of work, especially when affecting fundamental elements of the system. Due to this, the current design, which is still evolving, is foreseeing gradual migration to the new concepts without disrupting the normal flow of component production at large.

Re-factoring / fixing of existing code

As part of D4Science no large-scale development activities were planned, in terms of producing new functionalities and features. Nevertheless even at this stage, implementation was performed for reasons that become apparent below.

The second early-implementation objective was to cope with known issues of gCube that have been observed to compromise stability and availability of the system. Development in this direction was done at some services of the former Collective Layer (core services) and also some components of the Information Retrieval Services. Especially with regards to the Indices, new features that greatly enhance performance and stability have been obtained, and the gCube Result Set component (gRS) has been further optimized for performance.

After the official first release of gCore in March, significant effort was targeted on getting acquainted with the concepts, APIs and paradigms for implementation on top of the new framework, in order to perform the subsequent re-factoring of services. Initial adopters where services of the Core Layer which had to both support and exploit gCore: Information and Security Services were among these, yet not the only ones. Services in the Information Retrieval, such Data Source Description and Selection ones were also affected mostly as for proving concepts of the approach. The re-factoring of the rest of the services started before the end of the period and will be completed within the upcoming one, which implies that the main stream of re-factoring and fixing activities

products are expected to be made available before the first major deployment of gCube in the Environmental Monitoring Community (MSA1.1 Production Infrastructure at M6, MSA2.1a D4Science Portal Release and MNA5.1a ImpECt Virtual Research Environment must be delivered at M6).

However, re-factoring is not performed only in the direction of gCore adoption. Parallel to this, the teams have the opportunity to clean-up code and fine-tune it, which are both main tasks for D4Science, apart from delivering functional completeness to the user. Additionally, preparation of the code for accepting the foreseen functional changes has been performed in several cases, such as for instance the Search Operators and the Broker Service, as well as the already mentioned sub-systems of the Core Services.

Problems Encountered/Deviations from Plan and Corrective actions taken

Although no deviations or problems have occurred, there has been a need for proactive response to the following issue: the contact of new communities with gCube needs some time in order to produce user requirements. However, the tight plan of D4Science and the introduction of gCore platform as a unique instrument for the achievement of D4Science goals required that implementation starts almost immediately in the project. Yet, at the time of design for the delta implementation of gCube services in D4Science, user-requirements were not available. Without them, the risk of targeting the design away from user expectancies was raised. In order to buy the appropriate time, preliminary user requirements were gradually passed to the technical teams for early analysis and pre-planning of the JRA activities. In this way actual implementation will not deviate significantly from the results given by the full analysis of requirement, when the latter one is produced.

Achievements

- Requirements analysis and system design of gCore (3rd iteration);
- Advanced release of Index Service;
- Compilation of detailed dependency graph;
- Analysis of preliminary user requirements & compilation of system-oriented requirements;
- Overall planning of JRA activities;
- Delivery of tools for JRA activities;
- Reengineering of Core Services;
- Reengineering of Information Organisation Services;
- Reengineering of Information Retrieval Services;
- Reengineering of Presentation Services;
- Implementation and first internal release of gCore;
- Re-factoring of several Core Services for gCore adoption;
- Re-factoring of some Information Retrieval Services for gCore adoption;
- Issue-fixing in Information Retrieval Services and Components.

2.5 Standards Promotion and Adoption Activities Progresses

The D4Science consortium is deploying an e-infrastructure whose exploitation goes well beyond the groups represented in the project. This is intrinsically related with the scope of any infrastructure. However, being D4Science a Service Oriented Infrastructure it must be by definition open and capable to accommodate easily the needs and requests of unknown resource providers and consumers. This strong requirement becomes feasible only in the light of a strong adoption of standard approaches, procedures, and specifications as the ones prescribed by the gCube technology exploited by D4Science.

The list of standard specifications adopted by gCube is reported in the D4Science Quality Assurance Web Site accessible at:

https://quality.wiki.d4science.research-infrastructures.eu/quality/index.php/Standards.

It is important to note that the adoption of standard specifications does not raise the cost of participation to the infrastructure. On the contrary, the participation to the

D4Science infrastructure is strongly simplified thanks to the gCube Core, also called gCore, Distribution that embeds and makes transparent to the adopters many of the standard specifications needed to join it. Thus, the complexities associated with a cost-effective, secure, dynamic, and both short and long lived collaborations among remote parties is embedded in gCore that grants for free secure lifetime, scope, state, and configuration management of any resource joining the infrastructure. gCore does not propose a customized approach by creating a new proposal for standard. Rather, it offers an integrated framework that orchestrates and integrates by simplifying the existing standard specifications.

As a consequence, the participation to the D4Science infrastructure can be done either by following the gCore documentation that drives in the adoption of gCore or by complaining with the standard specifications recommend by OASIS, OGF, and W3 and their standard implementation released by Apache, Globus, and others well known projects.

3 PERFORMANCE INDICATORS

3.1 Coordination and Management Activities Indicators

Table 1 shows indicators capturing the activity of the D4Science project in the large. In particular, it shows the number of deliverable and milestones successfully delivered in the reporting period, the average deviation time with respect to the planned deadline, the average time requested to complete deliverable review process and figures about meetings and phone conferences organised.

Table 1. Project Coordination and Management Indicators

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Total
% Deliverables managed	7 / 7	/	/	/	/	/	/	/	7 / 7
Deliverables Completed	3								3
NA Deliverables Completed	1								1
SA Deliverables Completed	2								2
JRA Deliverables Completed	0								0
Deliverables Under Review	1								1
NA Deliverables Under Review	1								1
SA Deliverables Under Review	0								0
JRA Deliverables Under Review	0								0
Deliverables Ongoing	3								3
NA Deliverables Ongoing	2								2
SA Deliverables Ongoing	1								1
JRA Deliverables Ongoing	0								0
% Milestones managed	11 / 11	/	/	/	/	/	/	/	11 / 11
Milestones Completed	9								9
NA Milestones Completed	4								4
SA Milestones Completed	0								0
JRA Milestones Completed	5								5
Milestones Ongoing	2								2
NA Milestones Ongoing	2								2
SA Milestones Ongoing	0								0
JRA Milestones Ongoing	0								0
Avg Deadline deviation (day)	5.92								5.92
Deliverable	10.57								10.57
Milestone	1.27								1.27
Avg Deliverable review time (day)	5.85714								6
# Meetings	6								6
# Meeting Attendees	107								107
# Phone Conferences	8								8
# Phone Conference Attendees	51								51

This first reporting period is characterised by 7 deliverables and 11 milestones. All deliverables have successfully started but only 3 of the 7 planned have been sent to the European Commission while the other 4 are under internal revision. This status is partially justified by a set of reasons among which the legal issues related to the FAO participation was the most important and in fact FAO is directly involved in three of the missing deliverable. Moreover, the procedure put in place to guarantee the quality of the D4Science outcomes requested time to be appropriately tuned and perceived by the consortium members. However, thanks to the project support and monitoring tools, the governing board are well aware of the status of these deliverables and have put in place actions to boost the steps needed to complete them. With respect to the milestones, more than the 80% (9 out of 11) have been successfully completed. The two missing ones (MNA1.2 'Statement on the promotion of equality' and MNA4.1 'Training support environment') are ongoing and will be completed in the next period.

The indicators on the deviation time and review time indicate respectively that (*a*) the designed procedures, when adopted, guarantees the on-time delivery – except for the missed deliverable raising the average deviation time considerably; (*b*) the quality of the deliverables produced by the appointed editors is strong since the first version because the review time – including comments, comments reactions and validation – is relatively short (less than 6 days).

The reporting period has been also characterised by 6 project meetings and 8 phone conferences. Such networking activities involved more than 150 people. These events and their participation contributes to establish strong collaboration between the project members involved in the various activities in addition to the other collaborative tools deployed to serve the D4Science consortium operation.

3.2 Outreach Activities Indicators

3.2.1 Communication and Dissemination Indicators

Table 2 shows standard metrics capturing the impact of the Web site like the number of visits it receives, the number of visitors, or the average visit duration time.

Table 2. Web site impact

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Total
# visits	24346								24346
# visitors	4321								4321
Pageviews	60799								60799
Pages/Visit	2								2
Bounce Rate	0								0
Avg. Time (visits duration in sec)	312								312
% New Visits	0								0

Statistics for M1 are not included as the web site was published on 01/02/08.

Table 3 shows metrics on the dissemination events organised or participated as well as on the dissemination material distributed.

Table 3. Communication and Dissemination Activities Indicators

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Total
# Events	3								3
# Attendees	7								7
Dissemiantion Material Available	Flyer								
Distributed Material	308								308

D4Science flyer: 250 flyers were printed. 58 flyers were downloaded from the web site.

For what concerns the dissemination events, D4Science and its related topics have been presented at three main events (the 3rd Grid & e-Collaboration workshop for the Earth Science Community, the First DRIVER Summit and the 3rd EGEE User Forum). Each event was attended by at least two D4Science representatives presenting the project objectives and early findings. The potential audience reached by partaking to such events was significant in size (about 100 + 40 + 150 people), reference discipline (e.g. Institutional Repositories, Earth Science, Crystallography, Open Access, Finance & Multimedia), provenance and expertise.

3.2.2 Training Indicators

Table 4 shows metrics on the training events organised or participated as well as on the training material distributed.

Table 4. Training Activities Indicators

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Total
# Events	1								1
# Attendees	20								20
Training Material Available	Various								
Distributed Material	-								0

During this reporting period one official training event (a full day internal training) has been organised jointly with the 2nd D4Science TCom Meeting. The goal of the training event was to introduce D4Science members with the features of the new release of the gCore technology. Representatives of all technical partners attended the event and thanks to the supporting material prepared by the lecturers (slides, how-tos and code snippets) a pyramidal training model can be easily implemented (attendees can in turn train their colleagues that have not attended the event).

3.2.3 Communities VREs Definition, Validation and Exploitation Indicators

Table 5 contains various indicators on the activities of the user communities with respect to the definition and validation of the Virtual Research Environments.

Table 5. Communities VREs Definition, Validation and Exploitation

	Q1		Q2		Q3		Q4		Q5		Q6		Q7		Q8		
	EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	Total
# active VREs	1	n/a															1
# organisations involved	1	4															5
# people involved	2	8															10
# requirement items raised	5	0															5
# data sources identified	5	13															18
# data types identified	n/a	4															4
# tools identified	1	3															4
# requirements gathering meetings	0	4															4
# communities validating VRE	n/a	n/a															0
# people validating VRE	2	n/a															2
# bugs raised	n/a	n/a															0
# documentation errors	n/a	n/a															0
# changes to requirements	n/a	n/a															0
avg # disciplines involved in VRE	n/a	5															5
avg # organisations involved in VRE	1	4															2.5
avg # groups involved in VRE	1	6															3.5
avg # users involved in VRE	1	8															4.5
# report types	n/a	n/a															0
# report formats	n/a	n/a															0
number of accesses	n/a	n/a															0

In the first quarter 3 external organisations committed to participating in one or both VRE's as well as two FAO internal services. Requirements gathering has been an ongoing process, with both VRE scenarios underway. Tool identification has lagged behind data source identification, a situation that will be remedied in Q2.

3.3 Service Activities Indicators

3.3.1 Infrastructure Size

This set of indicators shows the number of resources constituting the D4Science production infrastructure and presents their evolution along the project lifetime.

Table 6. Production Infrastructure Size Indicators

		Q1		Q2		Q3		Q4		Q5		Q6		Q7		Q8		
		EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	Total
# Sites		4	-															4
# Active VOs		0	-															0
# Nodes		0	-															0
	# gCube nodes	0	-															0
	# Community nodes	0	-															0
	# gLite nodes	0	-															0
# gLite WNs		0	-															0
Storage (GB)		0	-															0
Availability	gCube nodes	0	-															0
	gLite nodes	0	-															0
# Data sources		0	-															0
# Data types		0	-															0
# Tools		0	-															0
Time-to-market		0	-															0

As described in section 2.3, during this first quarter, most of the effort was dedicated to the planning of the production infrastructure. Even if the initial preparation of sites has already started, the nodes are still not certified so cannot be counted.

3.3.2 Infrastructure Usage

This set of indicators shows metrics on the exploitation of the D4Science Production Infrastructure. Table 7 presents metrics on number of users, number of organisations served and number of Running Instances deployed per Community.

Table 7. Production Infrastructure Usage Indicators

	Q1		Q2		Q3		Q4		Q5		Q6		Q7		Q8		
	EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	Total
# Active users	0	-															0
Served organisations	0	-															0
# Active RIs	0	-															0

The production infrastructure is still being deployed. Therefore, no metrics can be collected about its usage.

3.3.3 Infrastructure Support

Table 8 presents the metrics on number of support request raised versus the production infrastructure for various reasons, ranging from malfunctioning to request for support, request for information, etc.

Table 8. Production Infrastructure Requests Support Service Indicators

	Q1		Q2		Q3		Q4		Q5		Q6		Q7		Q8		
	EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	EM	FRM	Total
Tickets opened	0	-															0
Tickets closed	0	-															0
Tickets duration time	0	-															0
Tickets duration steps	0	-															0
Escalated tickets	0	-															0

The production infrastructure is still being deployed. Therefore, no metrics can be collected about the support activity.

3.3.4 Software Release

Table 9 presents indicators capturing the software integration, testing and distribution activities.

Table 9. Software Integration, Testing and Distribution Indicators

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Total
# of components to integrate	18								18
Grow factor	-								-
Avg number of build executions per day	1.15								1.15
% of time head build is successful	18.00%								18.00%
Avg success ratio of integration and head builds	0								0
# build defects submited	0								0
# build defects closed	0								0
Avg deployment test success	0								0
# deployment test defects open	0								0
# deployment test defects closed	0								0
# functional test defects open	0								0
# functional test defects closed	0								0
Code quality (# potential bugs /1k LoC)	-								-
Coverage of documentation in code (# documentation lines / 1k LoC)	0.8249								0.8249
Coverage of documentation in manuals	-								-

As described in section 2.3, during this first quarter, most of the effort was dedicated to the planning of the integration, testing, and distribution procedures. No official release cycles have started. The results presented above refer to the gCube and gCore builds preformed over the code under development.

3.4 Joint Research Activities Indicators

Table 10 reports indicators capturing the software consolidation and development activities implemented to satisfy user requirements.

Table 10. Software Completeness

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Total
# Requirements satisfied	-								-
Time-to-market	-								-

Table 10 is not applicable in this reporting period, as the first release of user requirements is due to be announced after month 3 of the project (DNA5.1 Communities Practices and Requirements). Additionally since, minimal releases are just through the stage of re-factoring, the internal reviewing progress cannot validate statements on these indicators.

Table 11 complements the above indicators by presenting details on the partners' activities to develop the gCube software. Such activities will be measured in terms of number of components each partner is responsible for (also by type), the "size" such components have (expressed in terms of number of classes, Lines of Code and number of methods), the effort allocated (expressed in terms of number of developers actively involved in), the average deviation with respect to the development plan.

Table 11. Software Consolidation and Development Activity Indicators

	CNR	NKUA	ENG	BDM USTRATH	UNIBASEL	Total
# Components	18	62	13	4	16	113
Services	10	31	6	2	13	62
Libraries	4	13	7	2	2	28
Portlets	4	18	0	0	1	23
SW Size: #Classes	257	171	0	166	0	594
SW Size: LoC	14436	14453	0	8252	0	37141
SW Size: # methods	1383	1299	0	1065	0	3747
# Developers	7	7	2	1	2	19
Avg Deadline Deviation (days)	-	-	-	-	-	-

The numbers shown above are quite large, taking into account the short time elapsed since the beginning of the project. This is due to two facts:

- D4Science does a warm start from inherited DILIGENT produced system (gCube);
- gCube is quite large and extends in several layers of the eInfrastructure software stack.

Another fact is that the numbers are unstable due to the fact that the project is at its initial stage, and commits are performed by assigned partners not only on the basis of "ownership". Due to this, at this initial stage it is shown that BDM USTRATH numbers are quite high, i.e. quite higher than anticipated, due to the fact that the team of the organisation is also credited for external commits, a deviation which will be eliminated in the long run. Also ENG's numbers are misreported due to the fact that their work is part of the Core Layers which is not committed to the repositories by the same organisation.

Furthermore UNIBASEL's rate of commits is observed to deviate at the other edge due to internal work-team restructuring which is performed during this period.

More statistics:

Packages	Classes	Functions	NCSS	Javadocs	per
122.00	594.00	3747.00	37141.00	3064.00	Project
	4.87	30.71	304.43	25.11	Package
		6.31	62.53	5.16	Class
			9.91	0.82	Function

Average Function NCSS: 7.03
Average Function CCN: 2.87
Average Function JVDC: 0.63
Program NCSS: 37,141.00

4 NETWORKING EVENTS

Date	Activity	Location
07/01/08	gCube Core Meeting	Sophia Antipolis (FR)
08/01/08	D4Science 1st TCom	Sophia Antipolis (FR)
10/01/08	D4Science Kick-off	Sophia Antipolis (FR)
11/01/08	D4Science Quality Assurance Meeting	Sophia Antipolis (FR)
16/01/08	3rd GRID & e-Collaboration Workshop for Earth Science Community	Frastati, Rome (IT)
16/01/08	1st DRIVER Summit	Goettingen (DE)
21/01/08	D4Science NA3	Phone Conference
22/01/08	D4Science SA1	Phone Conference
13/02/08	EGEE User Forum 3	Clermont-Ferrand (FR)
20/02/08	D4Science SA3	Phone Conference
25/02/08	gCore TC	Phone Conference
28/02/08	D4Science PEB	Phone Conference
05/03/08	D4Science SA3	Phone Conference
06/03/08	D4Science user community meeting	Rome (IT)
12/03/08	D4Science SA1	Phone Conference
14/03/08	D4Science QATF	Prone Conference
17/03/08	D4Science 2nd TCom	Pisa (IT)
26/03/08	D4Science EAB	Pisa (IT)