PROJECT FINAL REPORT

Grant Agreement number:	**212488**
Project acronym:	**D4SCIENCE**
Project title:	**DIstributed colLaboratories Infrastructure on Grid ENabled Technology for Science**
Funding Scheme:	**CCPCSA (I3)**
Period covered:	**from 01/01/2008 to 31/12/2009**

Name of the scientific representative of the project's co-ordinator, Title and Organisation:

Philippe ROHOU, project coordinator, ERCIM

Tel: +33 4 97 15 53 06

Fax: +33 4 92 38 50 11

E-mail: philippe.rohou@ercim.eu

Project website address: http://www.d4science.eu/

Project deliverable: DNA1.6 **Submitted: 11/03/2010**

TABLE of CONTENTS

1. Final publishable summary report

Project Objectives

The main objectives of D4Science were to create a production quality e-Infrastructure for addressing the needs of two major target disciplines: *Environmental Monitoring* and *Fisheries and Aquaculture Resources Management*. This e-Infrastructure has been obtained by deploying, progressively enriching and consolidating the ones built by the EGEE and DILIGENT projects, and now it gives scientists of the addressed and other challenging areas more facilities for creating Virtual Research Environments (VREs) based on shared computation, storage, and generic service resources as well as on data and domain-specific service resources. Computation, storage, and generic service resources are offered by EGEE, project partners and their collaborators at a European level, while data and domain-specific service resources are offered by large international organizations, such as the European Space Agency, the Food and Agriculture Organization of the United Nations, and the Consultative Group on International Agriculture Research. The D4Science e-Infrastructure is expected to provide a multiplicative benefit to many scientific fields, and also act as a catalyst for the kind of cooperation and cross fertilization among multiple communities that is necessary for addressing many grand challenges of science and society.

Ten contractors and a sub-contractor participated in the project activities:

1. European Research Consortium for Informatics and Mathematics (ERCIM, France) *
2. Consiglio Nazionale delle Ricerche (CNR-ISTI, Italy)
3. National and Kapodestrian University of Athens (UoA, Greece)
4. European Organization for Nuclear Research (CERN, Switzerland)
5. Engineering Ingegneria Informatica SpA (ENG, Italy)
6. University of Strathclyde (USG, United Kingdom)
7. Universität Basel (UNIBAS, Switzerland)
8. European Space Agency (ESA-ESRIN, Italy)
9. International Center for Living Aquatic Resources (WorldFish Center, MY)
10. 4D SOFT Software Development Ltd. (4D-Soft, Hungary)

* The Food and Agriculture Organization of the United Nations (FAO, IT) is a sub-contractor of partner 1 ERCIM and plays a major role in the Dissemination and user Communities work packages.

The D4Science project management is jointly carried out by two individuals:

- Philippe Rohou - Administrative and Financial Coordinator, and contact point to the European Community
  ERCIM EEIG (European Research Consortium for Informatics and Mathematics, a European Economic Interest Grouping)
  2004 route des Lucioles - BP93
  F-06902 Sophia Antipolis Cedex, France

Tel: +33 4 9715 5306 e-Mail: Philippe.rohou@ercim.org

- Donatella Castelli – Technical and Scientific Co-ordinator
  CNR-ISTI (Consiglio Nazionale delle Ricerche - Istituto di Scienza e Tecnologie dell'Informazione)
  Area di Ricerca di Pisa
  Via Moruzzi, 1
  56124 Pisa, Italy
  Tel: +39 050 315 2902 e-Mail: Donatella.Castelli@isti.cnr.it

To succeed in its overall mission, the work performed in the two years of the project was articulated in the following five major areas of activities:

1. **Release and maintenance of the production-level e-Infrastructure**
2. **Set up of Virtual Research Environments (VREs) including application-specific content and services**
3. **Consolidation and Extension of the gCube system**
4. **Definition of strategies for outreach, exploitation, and sustainability**
5. **Scientific, Technological and Administrative Management**

Below a brief description of the objectives of each area and of its main achievements is given.

1. Release and maintenance of the production-level e-Infrastructure

This area of the project aimed at bringing the DILIGENT testbed infrastructure in production and at preserving its usage dependencies on the EGEE production infrastructure. This included the definition of procedures regulating the operation and enrichment of the infrastructure and the provision of appropriate tools for its monitoring and management. In the course of the project the developed e-Infrastructure was managed and upgraded by periodically deploying new community-specific resources as well as more consolidated and extended releases of the supporting system, i.e., gCube.

Main achievements

The D4Science production infrastructure has been deployed and made available to all project user communities as initial planned in the project DoW. An initial infrastructure milestone (MSA1.1) was reached in the summer of 2008 providing VREs to the Earth Monitoring (EM) community. The second infrastructure milestone (MSA1.2) was reached in the beginning of 2009. It introduced the support to the Fisheries and Aquaculture Resources Management (FARM) community. Finally, the third infrastructure milestone (MSA1.3) was reached in November 2009 by providing the two project communities (EM and FARM) with a number of VREs that contained new and revised functionality addressing the requirements expressed throughout the project. These achievements have followed the project initial plans of gradually enlarging the production infrastructure during the course of the project in terms of resources and users. The infrastructure is currently organized in the following VOs and VREs:

- EM VO:
  - Global Chlorophyll Monitoring (GVM) VRE

- Global Vegetation Monitoring (GCM) VRE
- FARM VO:
  - Fisheries Country Profiles Production System (FCPPS) VRE
  - Integrated Capture Information System (ICIS) VRE
  - AquaMaps VRE

Besides the VREs listed above, two other VREs (under EM and FARM VOs), have been created for demonstration and training proposes.

The infrastructure gathered thousands of resources of different types:

- Hardware Nodes: The infrastructure is composed by 53 physical machines, provided by five partners: CNR, ESA, FAO, NKUA, and UNIBASEL. These machines are exploited to run gCube nodes (60 gHN containers running different gCube services) and gLite nodes (59 worker nodes and several other gLite services). Figure 1shows the evolution of the hardware resources the infrastructure has been provided with since the beginning of the project (Q8 data is not final). The graph shows an increase of more than 100% since the first infrastructure milestone in Q2. It should also be mentioned the significant contribution of CNR site to the infrastructure (close to 50%). Finally, one reference is to be made to the visible grow verified between QR4 and QR5, corresponding to the availability of MSA1.2, when FARM user community was also provided with VREs.
- Data Collections:
  - The EM VO has 25 content collections and 69 metadata collections. These collections contain various types of material ranging from European Environmental Agency Reports to MERIS Level-3 products, Earth images and AATSR data sets;
  - The FARM VO has 15 content collections and 28 metadata collections. These collections contain various types of material ranging from AquaMaps products to fact sheets, time series graphs and country maps.
- Running Instances:
  - The EM VO is equipped with the number of components needed to guarantee the operation of the planned VREs. The approximate number of instances (including the enabling services) is 200;
  - The FARM VO is equipped with the number of components needed to guarantee the operation of the planned VREs. The approximate number of instances (including the enabling services) is 200.

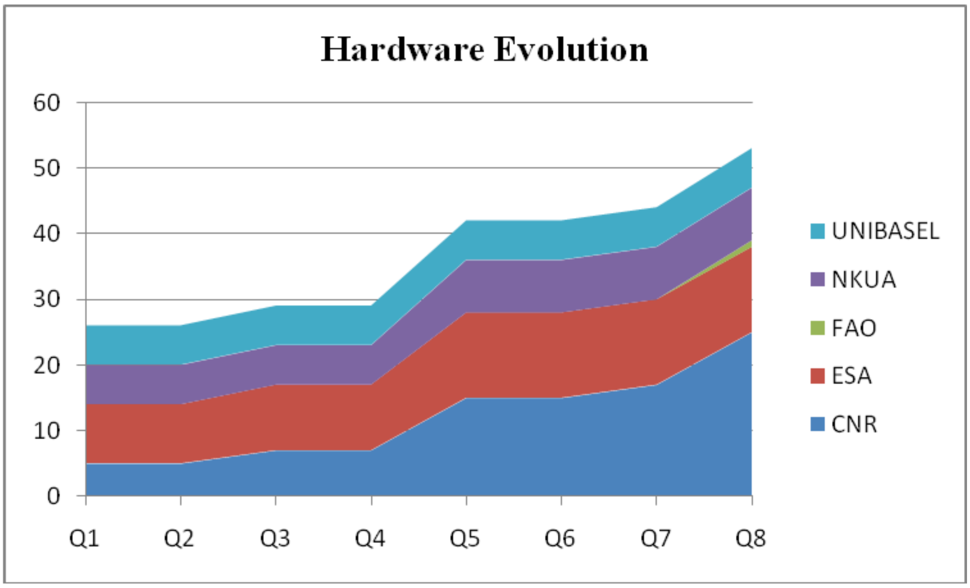

Figure 1 - Hardware resources evolution

- Data Collections:
  - The EM VO has 25 content collections and 69 metadata collections. These collections contain various types of material ranging from European Environmental Agency Reports to MERIS Level-3 products, Earth images and AATSR data sets;
  - The FARM VO has 15 content collections and 28 metadata collections. These collections contain various types of material ranging from AquaMaps products to fact sheets, time series graphs and country maps.
- Running Instances:
  - The EM VO is equipped with the number of components needed to guarantee the operation of the planned VREs. The approximate number of instances (including the enabling services) is 200;
  - The FARM VO is equipped with the number of components needed to guarantee the operation of the planned VREs. The approximate number of instances (including the enabling services) is 200.

The deployment of maintenance of VOs and VREs is based on clear procedures. These procedures are described in the Production Infrastructure wiki[1] while detailed information on how to put them in practice are described in the D4Science Administrator's Guide[2]. These procedures regulate the following two major activities:

VO Deployment:

- A VO enabling services deployment phase in which a VO Manager installs the mandatory services of the enabling layer needed to glue together the VO nodes;
- A VO level services deployment phase in which a VO Manager installs the set of services optionally deployed to support a specific deployment scenario;

[1] https://infrastructure.wiki.d4science.research-infrastructures.eu/production/

[2] https://technical.wiki.d4science.research-infrastructures.eu/documentation/index.php/Administrator's_Guide

- A VO configuration phase in which a VO Manager takes care of creating the conditions needed to instruct the VO enabling services and VO level services to implement the defined VO (e.g. preparing the service map, registering the mandatory resources);
- A VO management phase in which a VO Manager takes care of modifying the conditions governing the VO operation including the resources forming the VO.

VRE Deployment:

- A definition phase in which a user having the role of VRE Designer specifies the characteristics of a new VRE conceived to serve an application scenario;
- An approval phase in which a user having the role of VRE Manager decides whether the specified VREs have to be accepted or rejected. For what is concerned with the accepted VRE, the VRE Manager decides also how this VRE has to be deployed, e.g. which hosting nodes will be exploited;
- A verification phase in which a user having the role of VRE Manager validates a VRE resulting from the approval phase;
- A management phase in which a user having the role of VRE Manager operates on a deployed VRE in order to customise specific aspects (e.g. the layout governing the placement of user interfaces constituents) or monitor the operational state of the VRE as a whole.

The deployment and exploitation of these VOs and VREs is based on the gCube and gLite software. The infrastructure is currently based on gCube 1.5 and gLite 3.1 releases. All gLite nodes have been upgraded to release 3.1 when it was made available by the EGEE project. Afterwards all released patches have been applied. Concerning the gCube nodes, six main releases have been integrated, tested, documented, and deployed by the project Service Activity: gCube releases 1.0 to 1.5. For each release, a number of different maintenance releases were also produced in order to fix critical defects:

	Upgrade Completed	Upgrade Duration	Maintenance Releases	Milestone
r1.0	22/10/08	-	0	MSA1.1
r1.1	09/03/09	-	4	MSA1.2
r1.2	31/07/09	240	3	MSA1.2
r1.3	01/10/09	69	0	MSA1.2
r1.4	06/11/09	96	0	MSA1.2
r1.5	20/11/09	10	0	MSA1.3

Table 1 - gCube releases

Whenever a new release was completed and made available for deployment in production, an infrastructure upgrade plan was defined by the Infrastructure Manager. Such plan defines the nodes to be upgraded, by whom, and when.

shows the evolution of the number of hours needed to upgrade the infrastructure. While the first infrastructure upgrade took weeks, the most recent upgrades were carried out in few hours. This is a direct consequence of several improvement introduced in gCube related to infrastructure operation.

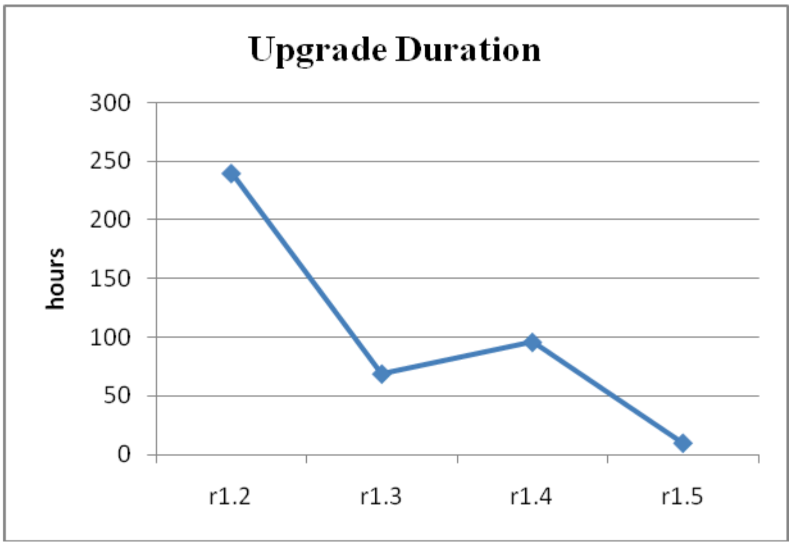

Figure 2- gCube nodes upgrade duration

As for the infrastructure uptime, a total of 19 gCube downtimes and 32 gLite downtimes were declared by Site Managers since the initial deployment of the infrastructure services. It should be mentioned that downtimes were mainly caused by either scheduled network interventions or major infrastructure upgrades. Most of the downtimes did not last more than one day so the production infrastructure was never affected by long unavailability periods.

The monitoring and accounting of the D4Science production infrastructure are central activities within the SA1 Infrastructure Operation work package. Such activities were based on a number of tools developed or adopted by the project. Since the beginning of the project these tools have been gradually deployed in the infrastructure. Due to their distinct nature they can be grouped in two categories:

- Based on information collected by the gCube Information System (IS);
- Based on information produced/consumed by the gCube Messaging System (MS).

While the gCube Information system provides only tools for monitoring, the gCube Messaging System provides tools for monitoring and accounting. These tools provided large amounts of relevant information that allowed different infrastructures roles (Site Managers, Data Managers, etc) to visualize the status of their resources, to be actively notified when problems occurred, and to access statistics about users' exploitation and service-to-service communications.

A clear production support procedure was also defined, put in production, and revised during the course of the project. This procedure ensured an efficient response to all defects found in the production environment. A total number of 387 tickets were submitted, 42% being of high priority. All tickets were properly closed and documented. Figure 3 presents the distribution of tickets through the project lifetime and the percentage of high priority tickets.

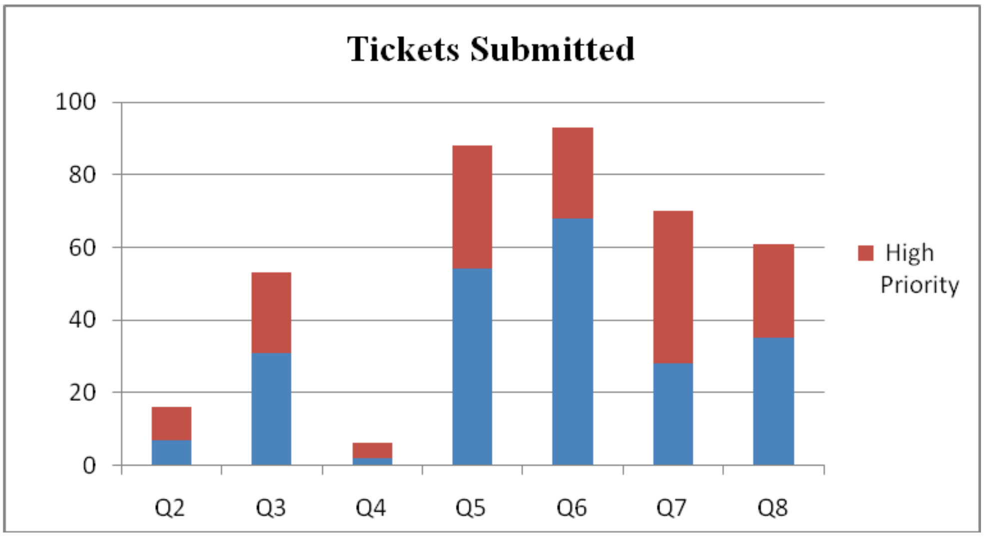

Figure 3- Production support tickets submitted

Finally, two other activities related to the production Infrastructure should also be reported:

- Two training events have been organized by the training work-package with direct support by the project Service Activity. The goal of these training events was to officially transfer the management of the D4Science production infrastructure from the project technical partners to the project user communities. The first event focused on the VRE-level aspects while the second touched the management of the infrastructure at ROOT and VO level.
- A collaboration with the Academia Sinica Grid Computing (ASGC) in Taiwan has been established. ASGC is particularly interested in the work carried on the AquaMaps VRE. ASGC is working to support the D4science VO in their gLite site. Afterwards, ASGC may also run a number of gCube nodes hosting services for the AquaMaps VRE.

2. Set up of Virtual Research Environments (VREs) including application-specific content and services

This area is concerned with the building of VREs intended to address the needs of target scenarios established by the scientific communities interested. It also deals with the adaptation of community-specific data and service resources included in the VREs in order to exploit the e-Infrastructure capabilities.

Main achievements

Five VREs (FCPPS, ICIS, AquaMaps, GCM and GVM), described below, were created and are currently maintained.

Requirements for these VREs were collected by community representatives. In particular, the FCPPS requirements were specified by FAO in a vision document evolving from a series of brainstorming meetings lead by FAO/FI top level management, towards identification of a new business process for the elaboration of Fishery Country Profiles. From the vision document, specific requirements were added to the trac system. These requirements were refined at several project TCom meetings. Progressively, they have also been enriched by the FAO User Community with the aim of adding support for Life-Cycle Management of documents, output to advanced XML schemas, concurrent editing, etc. The functionality of the FCPPS VRE is validated by the UC.

For ICIS, an inception meeting was convened in July 2008 to introduce the e-Infrastructure technology to potential users. The outcome of this meeting was a major re-adjustment in D4Science objectives, as the requirements for ICIS specifically aimed at solving problems in the statistical realm that are difficult to tackle using standard software and require functionality that has never been realized on other platforms.

These issues were further discussed in two design meetings, respectively in January 2009 and in July 2009, in Rome, focussing on Timeseries management in an e-Infrastructure. Here, experts from FAO and CNR discussed the implementation of a Timeseries management database in D4Science using an approach to data-management radically different from the one hitherto used. The core functionality has recently been released and validated. The User Community has seized the opportunity to submit further specify requirements, e.g. on mapping between data-sources and implementation of SDMX as a communication standard.

The AquaMaps VRE, not initially planned in the DoW, followed a different path, mainly because the requirements were very clear from the start. Its main objective was to port an existing complex and computing intensive algorithm to a grid based e-Infrastructure. CNR invited representatives from World Fish centre and FAO in May 2009 to discuss the parts of the existing AquaMaps application that would most benefit from the opportunities offered by the e-Infrastructure: i.e., computing power and support for geospatial data and services. The Project team decided to take a step by step approach, and focus on the generation of species distribution maps. This part of the VRE has been delivered and has been presented at several international meetings. Current developments are targeting geospatial functionality, interactive mapping and synchronization with external data sources.

During TCom meetings the D4Science e-Infrastructure capabilities were discussed with the user communities in order to facilitate their co-operation and release of data-sources. This consultation also gave the user communities the opportunity to further specify VRE requirements and adjust VRE development plans. These adjustments were discussed with other work packages, which resulted in adjustments in the scope and purpose of the proposed VREs. In particular, the User Communities identified a clear niche for the VREs minimizing overlap with standing software and procedures, and maximizing benefits for a potentially large audience. This resulted in some high level objectives for the already specified VRE scenarios, in order that:

- FCPPS provides an integrated reporting environment across VREs;
- ICIS provides a database for statistical data and their curation and modification;
- AquaMaps develops support mechanisms for geo-data;
- ESA's GCM and GVM provide an integrated reporting facility linked to on-demand processing environment.

In line with the ongoing nature of the VRE improvements, the consultation has been continued for the whole project lifetime with potential users, such as:

- for FCPPS: FAO Fisheries and Aquaculture Department units, and relevant FAO corporate units
- for ICIS: FAO, ESA, WorldFish, the Northwest Atlantic Fisheries Organisation (NAFO) and International Council for the Exploration of the Seas (ICES) and
- for AquaMaps: the ASEAN Center for Biodiversity, IUCN, GOBI (Global Ocean Biodiversity Initiative) as well as a large number of conservation NGOs and researchers

Collected requirements have been analysed, selected, and then transformed to the granularity of tickets in the TRAC system.
A strategy for VRE validation has been developed in close co-operation with the service activity. User support and documentation are now provided through a community WIKI.

In the following the five VREs mentioned above are described.

A. *Fishery Country Profiles Production System (FCPPS)*

The design of this VRE was led by requirements specified by the FAO Fisheries and Aquaculture department. It supports the standardized generation of fisheries and aquaculture reports containing information in country-specific formats which enhance decision-making and promote advocacy in the sustainable use and conservation of aquatic resources. The production of country profiles requires complex aggregation and editing of continuously evolving multi-lingual data from heterogeneous sources (see *Figure 4*). The FCPPS Virtual Research Environment was introduced to support the entire reports' lifecycle from data-collection, through template and content editing, co-authoring, reviewing and quality control until reports are ready for publishing. The VRE workspace provides dynamic links to data-sources, and provides the community with up-to-date information to include in the reports. An important feature is the integration of data across VREs, where e.g. data from EO or AquaMaps VREs are seamlessly integrated in a report, which would be much more difficult in other products.

The VRE also offers facilities for the introduction of new data sources that contain entire collections of information, provides a secure environment to discover, share and modify sometimes confidential information, and assists in structuring document. An advanced annotation system enables editors to quickly focus their attention on those sections that require editing. The design of the VRE and the supporting infrastructure allows the FCPPS community to focus on the reporting life-cycle, and drastically reduces the effort of managing large data-sets, multiple users and versions of reports throughout their life-cycle. The VRE provides storage of reports in baskets, together with information that may be required once an update is needed, and allows for publication in a variety of formats that FAO can use to convert the information to its specific web-publishing standards. The experience gained in the development of the FCPPS will be valuable not only for the other VREs and VOs, but also future VREs will benefit from the generic reporting functionalities.

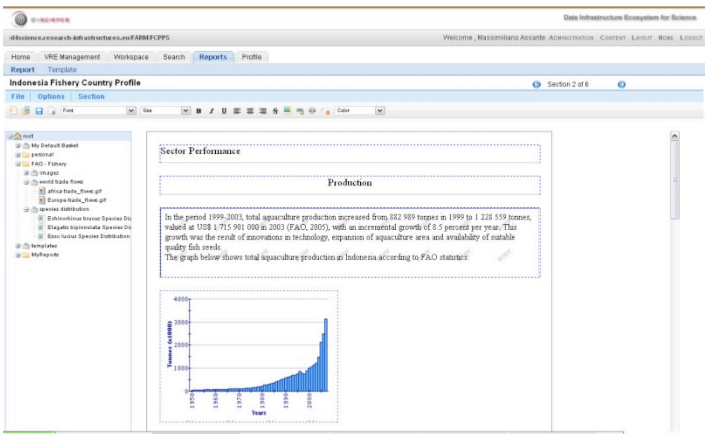

Figure 4 - The FCCPS VRE for report creation

B. Integrated Capture Information System (ICIS)

Also the design of this VRE has been led by requirements specified by the FAO Fisheries and Aquaculture department. The ICIS VRE focuses on integration of regional and global capture statistics of aquatic species, and manages both the data-sets and their metadata, as well as the complex relationships between these metadata coming from a number of Regional Fishery Management Organisations (RFMOs) and international organisations (FAO, WorldFish Center). It also supports tools for curation while importing these data-sets as well as the process from data import until the generation of standardized Timeseries datasets. This requires the configuration of algorithms and filters for parameters such as area, species distribution and habitat. These functionalities provide a harmonized view of catch statistics, thus allowing the community to overlay these statistics according to pre-defined reallocation rules. To ensure broad interoperability the VRE implements de-facto international standards for describing code lists, including those agreed at the Coordinating Working Party (CWP) on Fishery Statistics. The datasets and code-lists are prepared for compatibility with other standards, such as the open geospatial consortium's ISO 9115 and those for SDMX. This VRE is in part a response to the increasing demand for global catch data with finer geographical resolution, including:

- The United Nations recommendation to FAO to "revise its global fisheries statistics database to provide information for the stocks to which the Agreement applies, as well as to high seas discrete stocks on the basis of where the catch was taken." (Document A/CONF/.210/2006/15, Annex, Paragraph 25, 26);
- Requests to differentiate catch from Exclusive Economic Zones (EEZs) and high seas;
- The CWP recommendation to respond to the UN recommendation by integrating catch statistics that Regional Fishery Bodies (RFBs) and FAO have collected, maintained, and disseminated, based on rules established in the previous CWPs (CWP-22 Report).

These recommendations require the development of many features also outside the e-Infrastructure. A first step was the prioritization of requirements relevant to D4Science. The data marshalling and management requirements were developed first, and ICIS now offers fisheries statisticians a set of tools to manage their data. Statisticians produce statistics from often very different data sources, and the VRE offers control of the processes to import, manage, compare and export statistical data for fisheries (see Figura 5)..

The ICIS long term vision considers the D4Science data e-Infrastructure particularly suited to:

1. collect and analyze capture data;
2. check and control data and data-quality;
3. produce uniform data-sets adhering to standards such as SDMX;
4. provide a harmonized data-repository for fisheries data.

In the short to medium term plan, the standing activities of data collection, analysis and dissemination are left to existing systems in support to statistical units work (e.g. FIStatWS and FishStat+ in FAO), and emphasis for D4Science is placed on the ability to import,

integrate and most importantly harmonize and compare already published regional and global statistical sources.
In addition, the D4Science grid e-Infrastructure facilitates to share data across VREs and use statistical data in other use-scenarios, such as spatial analysis or reporting. The collaborative and computing environment also promises to ease the difficult transformations from one reporting framework to another. The VRE has made the process of producing statistics for e.g. FCPPS Country Reports much more efficient and reliable.

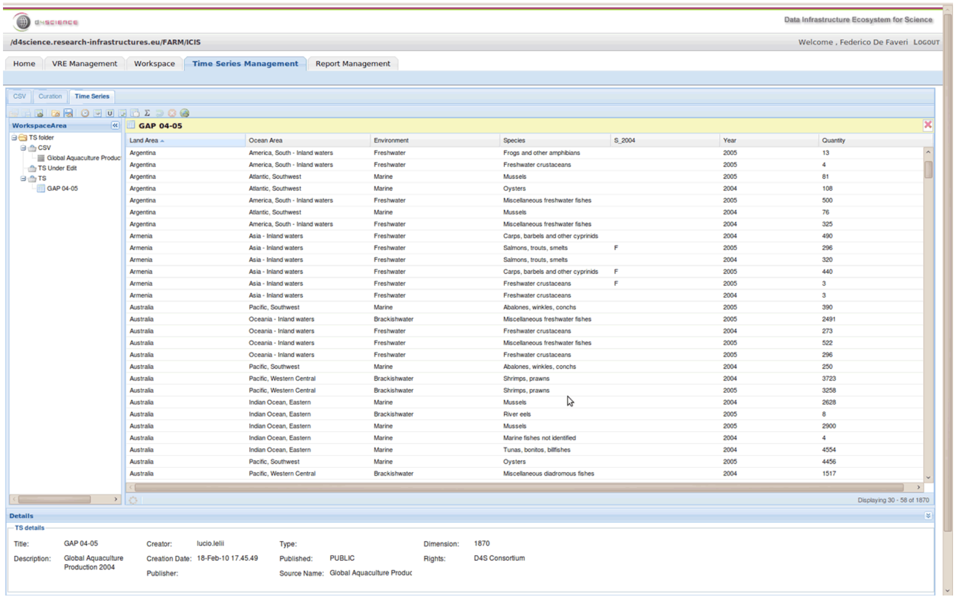

Figura 5- ICIS VRE for timeseries management

C. AquaMaps

The AquaMaps VRE is based on an existing species distribution modelling approach and associated data developed by the AquaMaps project (www.aquamaps.org) which aims to produce standardized range maps of (eventually) all marine species. The decision to incorporate and support this scenario in D4Science was taken during the course of the project. Although it was apparent from the start that it would require additional effort, it was selected because it can show the potential of the D4Science infrastructure. Moreover, the resources produced by this scenario are also useful in the FCPPS and other VREs related to the biodiversity domain. The requirements for this scenario were provided by the WorldFish Center, FAO and the FishBase Information & Research Group.
The AquaMaps Virtual Research Environment (Figure 6) is designed to provide fisheries and aquaculture scientists with a collection of species distribution maps as well as the underlying matrices of the relative probability of occurrences. Within the AquaMaps VRE predictions about species distributions can be generated for a large number of species across all major oceans and continents. These already widely used maps are an important tool in species assessment.
The VRE was also designed to better understand the tools and standards required to produce quality geo-data catering for the ever increasing demand for geo-referenced products. The AquaMaps collections have been produced by processing data of 9154 species and 56,468,301 database records representing the relative probability of occurrence of a given species per half degree latitude by longitude surface grid cells. Five collections have now

been generated by filtering, grouping and rendering these data: species maps, family maps, order maps, class maps and phylum maps. These collections contain compound objects that consist of images and their related metadata. The produced maps are available in several geographic projections, including standard global maps but also stereo-polar views centering on the poles or maps focusing on specific other areas etc.

The AquaMaps VRE is not a static environment, and it has raised considerable interest for further development. In the future, the VRE could evolve into a flexible and standards based geospatial tool. The currently implemented algorithm, for instance, is only one of many methods available to scientists to analyze and predict species distribution. There are other datasets on environmental and species preferences to consider. Also, the maintenance of provenance of species occurrences is important, and the synchronization of VRE data with source data can benefit from the data-management capabilities of the VRE. The processing of geospatial data requires computing power that can be delivered by the e-Infrastructure. More information on AquaMaps is available in the video "A Virtual Research Environment for Species Distribution Map Generation and Management", http://www.youtube.com/watch?v=JhPISI-9O9A

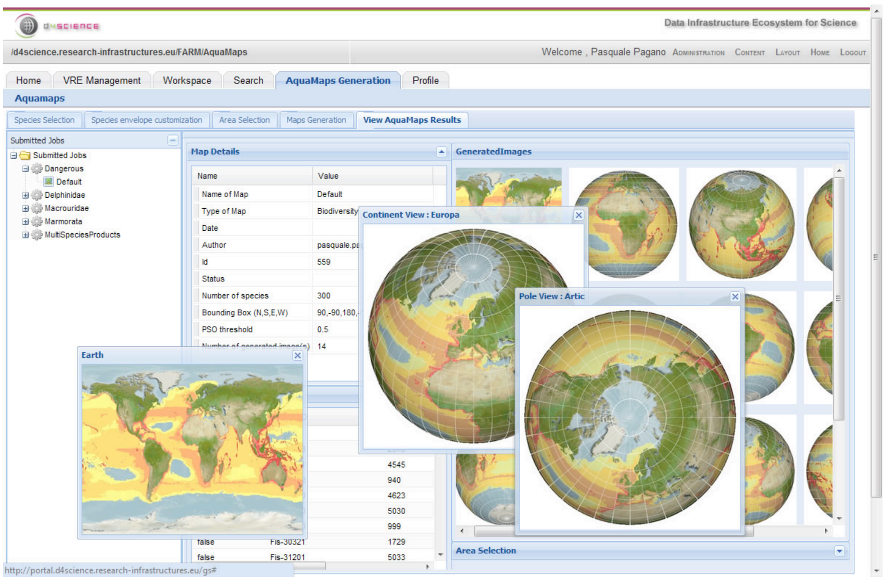

Figure 6 - AquaMaps VRE for species map generation

D. Global Ocean Chlorophyll and Land Vegetation Monitoring (GCM and GVM VREs)

The requirements for this scenario have been provided by the ESA community which has also participated in and completed the validation.

The Environmental Monitoring scenario is based on the outcome of the experimentation performed in the framework of DILIGENT aimed at providing e-Infrastructure services to ease the task of implementing environmental conventions. Though the conceptual framework remains the same, starting from that scenario (environmental conventions implementation in support of climate change), user requirements have been expanded during D4Science targeting user friendly means to set-up on-demand and maintain user workspaces (VREs) enabling collaborative generation and maintenance of periodic environmental reports, on-demand dataset computation (data generation, data transformation), shared annotations, enhanced search facilities, easy plug-in and demission of relevant information sources.

To demonstrate and validate the system capabilities the community designed two VREs dedicated to the monitoring of global ocean chlorophyll concentration and land vegetation, with the intention of experimenting by themselves the VRE set-up, maintenance and use. To enable successful and timely transfer and adoption of the developed technology, ESA currently hosts the whole set of e-Infrastructure services produced with D4Science.

Both the VREs, named Global Ocean Chlorophyll Monitoring (GCM) and Global Land Vegetation Monitoring (GVM), aim at providing integrated and collaborative workspaces where scientists can retrieve, visualise and access heterogeneous information about the Earth, annotate them, share newly generated information via ad-hoc processing and create on-line documents, exportable in various formats (see Figure 7).

Conceptually, both the above VREs claim to provide support for accomplishing research tasks related to global environmental monitoring and climate change analysis; currently they cover two broad areas of Earth Science and host cross domains data and services.

The information currently available from both GCM and GVM VREs is quite heterogeneous and varies from satellite imagery, Earth observation (EO) products, technical notes and manuals to dissemination documents, reports and user artefacts. Particularly useful to the average user is the visualisation of entire EO datasets through the D4Science information object model: a single product can be visualised as part of weekly/monthly or yearly series, with as many metadata as required by the user, and can be retrieved on-demand from the ESA archives. Currently GVM can be configured to link with the ESA G-POD service for computing global vegetation indexes: the GVM user can submit computation tasks to the ESA GRID infrastructure, publish back the result in GVM and share it with GVM colleagues. The ESA team could also experiment with the D4Science technology from the administrative point of view, e.g. via configuring VREs, making information available, setting-up remote connection to e-Infrastructure services for searching and transforming data, and managing users. In support of VREs administration tasks, the project delivered monitoring tools that ease the exploitation of the platform as a ‘vertical’ one, from infrastructure enabling services up to the application level.

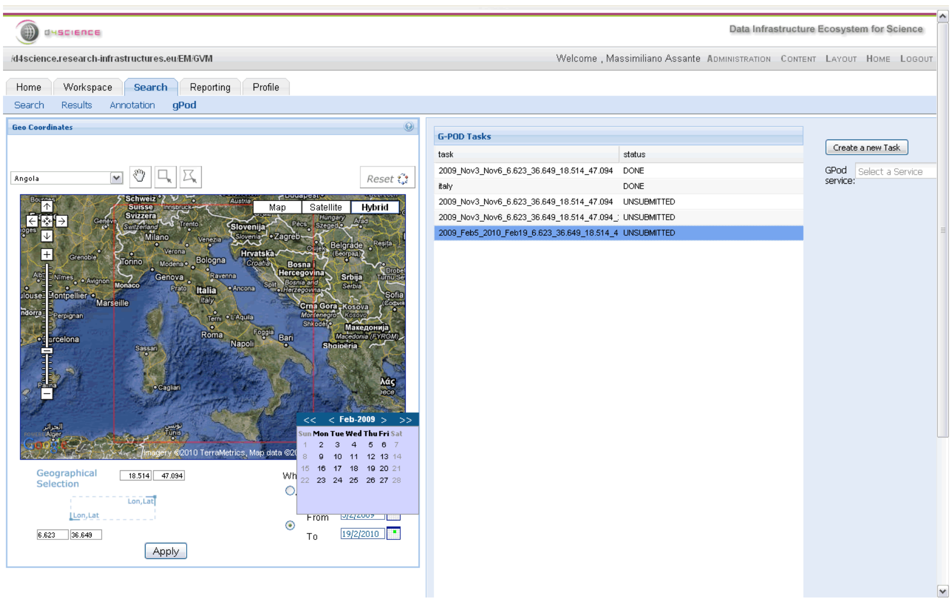

Figure 7 - Global Ocean Chlorophyll Monitoring VRE

3. Consolidation and Extension of the gCube system

This area focused on the continuous upgrading and extension of the gCube system. These activities were performed in order to offer better quality services and to meet particular specific requirements identified by the user communities but not yet covered by the current existing releases.

Main achievements

gCube is the D4Science enabling "software system". Its development started in the DILIGENT project following a standard software development methodology: requirements-design-implementation-verification-maintenance. In D4Science this methodology was changed in favor of an Agile software development methodology. This implies that new software components, or revised version of existing software components, have been developed in relatively short subsequent iterations, each containing the tasks needed to release the new functionality - planning, requirement analysis, detailed design, implementation, integration, testing and documentation. As a consequence of this methodology, a number of subsequent versions of the same component have been released during the project lifetime.

gCube is intrinsically complex because of the amount of standards, specifications, technologies, patterns and approaches put in place to guarantee a high quality software system (reliable, secure, autonomic, fault resilient, etc.), powerful enough to support the development and operation of Service Oriented Infrastructures enabling the building of Virtual Research Environments. In such development context, there was a severe risk that this complexity could overwhelm developers and distract them from the specific application logic of the services they were willing to develop.

In order to prevent the above risk, the D4Science Technical team decided, in the first months of the project, to deviate from the planned workplan and introduce the gCore Application Framework (gCF). The gCore Application Framework allows gCube services to abstract over functionality that are lower in the web services stack (WSRF, WS Notification, WS Addressing, etc.) and to build on top of them advanced features for the management of state, scope, events, security, configuration, fault, lifecycle, publication and discovery.

Thus, today, the D4Science enabling technology actually consists of two systems - gCube and gCore - each evolving by its own.

As far as gCore is concerned, it is a Java-based application framework for the development of gCube services and their clients. By design its goals are: (a) to simplify and standardize all systemic aspects of service development, particularly those which relate to the fulfillment of gCube-specific requirements; (b) to promote the adoption of best practices in distributed programming for concerns such as safety of concurrent access and autonomicity of behaviour. With its latest release (version 0.6) the framework:

- manages the entire *lifecycle* of gCube services, engaging in autonomic interactions with the infrastructure and the local environment, and allowing customisations on state transitions, including deployment, initialization, activation, and failure;
- enforces the *scoping* and *security* rules associated with shareable resources, handling the acquisition and renewal of service credentials, the delegation of caller credentials, and the propagation of scope and credentials from incoming to outgoing service calls;

- implements WSRF standards for publication, access, and notification of change in service *state*, offering a rich set of abstractions for modeling it, governing transparently its full lifetime, and managing its persistence on different storage media, including its recovery from remote media upon service migrations;
- standardises the use of systemic *faults* within service interfaces and implementations, transparently supporting retry-same and retry-equivalent semantics and converting faults into equivalent lighter-weight exceptions at service boundaries;
- mediates access to configuration resources on classpath and local file system, redirecting read failures to backups created prior to write operations and exposing object bindings for all aspects of service and container configuration;
- simplifies *resource discovery* via object bindings, templating, and XPath inspection for a range of queries to the Information services of the infrastructure;
- simplifies *multiprogramming* via arbitrary combinations of event-based synchronisation, scheduling, parallelisation, and sequencing of local processes;
- simplifies *distributed programming* through the customisation of best-effort discovery and interaction strategies with stateless and stateful services;
- includes programming abstractions for the design of *pluggable services*, i.e. services that can be dynamically extended with code that is packaged and deployed as the payload of *service plugins.*

As far as *gCube* is concerned, it is a software system that has been designed by following the principles of component-based software engineering, thus it consists of various software components interacting each other to implement the expected functionality. In particular, such a software system is built by combining in a Service Oriented Architecture a number of subsystems[3] supporting the following facilities:

- *Infrastructure Operation and VRE Management*. It includes (*i*) organisation and execution of Virtual Research Environments by guaranteeing an optimal consumption of the available resources; (*ii*) registration of the infrastructure constituents; (*iii*) authentication and authorization policy enforcement enabling the highly controlled sharing of infrastructure constituents; and (*iv*) definition and orchestration of complex workflows by guaranteeing an optimal consumption of the available resources. With respect to *VRE Management*, it includes the definition of VREs and the dynamic deployment of VRE resources across the infrastructure.

  VREs definitions are declaratively specified through an appropriate and user-friendly user interface in a dedicated language which is resolved by deriving an optimal deployment plan. The plan is based on availability, QoS requirements, resource inter-dependencies, and VRE sharing policies, but also on monitoring of failures (resources are dynamically redeployed) and load (resources are dynamically replicated). Three distinguished services (*Software Repository*, *Deployer*, *gCube Hosting Node Manager*) support VRE definition and dynamic deployment by, respectively, collecting service implementations, deploying service implementations and their dependencies on gHN, and hosting such service implementations at selected nodes.
- *Information Organisation*. It includes: (i) the storage and organisation of Information Objects (compound objects comprising multiple parts and alternative representations) and their constituents including the management of statistical data and metadata objects like time series; (ii) the management of the metadata objects in multiple formats equipping

[3] A subsystem is intended as a logical constituent unit when considered with respect to the system as a whole. A subsystem groups Services, Libraries and any other kind of software component belonging to the area of competence of the subsystem.

each Information Object; and (iii) the management of the annotations objects potentially enriching each Information Objects. In particular, a completely new subsystem dedicated to support the entire lifecycle governing statistical data, like time series, i.e. sequences of data points or measurements on a certain phenomenon at certain time periods, has been developed. It relies on the other Information Organisation services to create proper Information Objects, however, at the same time, it implements the necessary type specific application logic. In particular, the Time Series Management services support data import, data curation and data usage.

- *Information Retrieval*. It includes a comprehensive set of Information Retrieval (IR) facilities allowing searching over data and information by a wide range of techniques. It comprises (i) a search framework comprising a *search orchestrator component*, a set of *search operators*, a *query processor component* and a *data transfer mechanism*, (ii) an index management framework supporting all aspects of an index lifecycle as well as support for search capabilities for a rich set of indices, such as full text, forward, feature, geospatial indices; and (iii) a distributed framework implementing higher-level IR capabilities which include content ranking, source selection and result set fusion (ranked merging of various data sets).
- *Presentation Services*. Realises the logical top layer of a gCube-powered infrastructure with a twofold objective: (i) provide the means to build user interfaces for interacting with and exploiting the gCube system and infrastructure and (ii) provide a full range of user interfaces for achieving interaction with the system, out-of-the-box. In particular, a comprehensive set of user interface constituents have been developed to provide the gCube end-users with an attractive and easy to use front-end making the exploitation of complex facilities an easy task. Examples of front-end constituents are the *AquaMaps portlet* – a console providing its users with facilities for species occurrence maps production and comparison; the *Time Series portlet* – a console providing its users with an array of Time Series management facilities ranging from import to curation and filtering; the *gPod portlet* – an environment providing its users with facilities to interact with the ESA gPod service; and the Resources Management portlet – an environment designed and implemented to systematise the facilities for “resources” management previously spread in multiple portlets.

4. Definition of strategies for outreach, exploitation, and sustainability

This area was devoted to define strategies for exploitation, sustainability, and outreach. These strategies had to be directed not only at the two communities involved in the project but also at other scientific communities with similar requirements. The outreach objectives included, among the others, offering training, as well as consulting and technological support, to those communities that plan to register resources and use the functionality of the D4Science e-Infrastructure.

Main achievements

The main outreach activities were devoted to diffuse information and raise awareness about the D4Science project in the user communities; promote participation and further strengthen ties with the user communities; encourage the e-Infrastructure to be a tool assisting the work of scientists; establish links and encouraging synergies with similar projects and initiatives; and raise new synergies and collaboration opportunities among consortium members. They also offered training, as well as consulting and technological support, to those communities that plan to register resources and use the functionality of the D4Science e-Infrastructure.

The communication strategy applied encompassed the three main steps of dissemination, i.e., Awareness (Web site and Internet presence, Newsletter – Magazine, External events, Press releases, E-mailing, Link exchange, Publicity material), Understanding (Publications Dialogue, debate and networking, Sharing of standards), and Action/Participation (D4Science portal, Case studies and best practices, Feedback mechanisms, Training).

The Outreach activities covering the three steps of dissemination above, i.e., awareness, understanding and action/participation, evolved over the course of the project.

The initial period was dedicated to defining the communication strategy, aimed to vehicle the project's key messages to the scientific user communities.

The D4Science project's web site (www.d4science.eu, Figure 8), a core tool to disseminate project's results, was published online in the first months of the project, (May 2008) and it is still active, since the project D4Science II has decided to keep it. The project's web presence was increased via the creation of wikis, portals, the organization of link exchanges with related web sites, the addition of project descriptions on partners' web sites, etc.

Since then, the web site content was enriched, new features were added and some Web 2.0 technologies used to ease and broaden access to project's resources and information. Regular posting of news, events, dissemination material, papers and deliverables kept the audience informed of new and ongoing project activities. Monthly traffic weblogs are analysed to identify visits, visitors, referrers and also potential access issues.

Unique visitors	Number of visits	Pages
71,550	239,186	1,118,094

D4Science web site logs – selected numbers (Jan 08 – Dec 09)

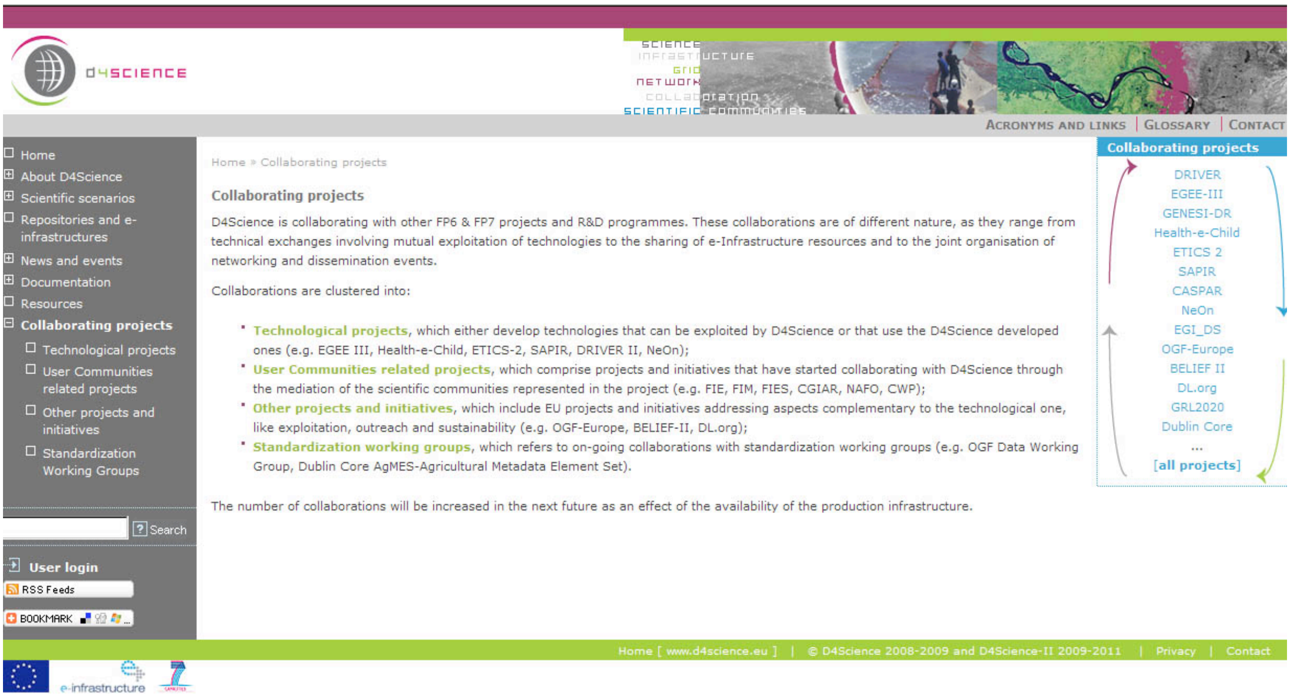

Figure 8 - The project Website

The collaboration with other projects was strongly highlighted through a dedicated section on the web site. With the preparation of articles D4Science contributed to newsletters, bulletins, magazines and joint publications. Many contacts were established via e-mail but also via face-to-face meetings.

Project's members organized and/or participated in about 30 selected dissemination events, some particularly orientated to the user communities and potential scientific users, others addressing technological/standards aspects (ECDL, VLDL, AHM, OGF, etc.). One of the key achievements of the Communication and Dissemination activities was the organization of the World User Meeting that was held at The Food and Agriculture Organization of the United Nations (FAO) in Rome, Italy, 25-26 November 2009. More than 80 experts were present from about 20 countries (Figura 9). In total, in less than 2 years, D4Science members physically presented D4Science or certain aspects of it to over 2000 persons.

Furthermore about 20 training sessions took place attended by a total of 290 participants.

Figura 9 – D4Science World User Meeting participants

Other dissemination activities included slides presentations, organisation of and participation in sessions and demonstrations. On such occasions the crucial collaboration with other projects (EGEE, BELIEF II, DELOS, GRL 2020, CASPAR, DRIVER, DL.ORG, DARIAH, etc.) or communities related initiatives (IMDIS, COFI, CWP, IODE ,OBIS, FishBase, etc.) was also reinforced.

Dissemination material, ranging from slides, articles, press releases, posters, brochures and videos, were prepared for and distributed at events (Figure 10).

Figure 10 - D4Science posters

5. Scientific, Technological and Administrative Management

This last area concerned the overall management of the project from the scientific, technological and administrative point of view. In particular, it comprised the definition of a detailed workplan, the monitoring and assessment of the progresses of the project with respect to this workplan, the measurement of the quality of the activities performed and the prompt application of corrective actions. It also included a careful monitoring of the resources and financial expenditures and the fulfilment of contractual obligations and reporting.

Main achievements

A considerable effort was dedicated to manage the project from the scientific, technical and financial point of view. Activities were continuously monitored from these three different perspectives in order to assure appropriate quality and progresses. To help this process, a Quality Assurance Task Force, appointed at the beginning of the project, defined a project quality plan and monitored it. The plan defines procedures concerning a considerable number of project aspects such as Project Governing Bodies, Meeting Procedures, Risk Management, Conflict Resolution, Activity Reporting, Deliverables Management, Software Release, Infrastructure Operation, etc. Most of this information is available in a wiki website dedicated to the project quality aspects. Many other tools, like the wiki which maintains the status of the deliverable workflow process, have been introduced to support managerial activities. Thanks to these tools, activities, like the monitoring of the project status, the planning of future activities and the identification of corrective actions, are largely simplified.

The production workflow of each of the 5 deliverables submitted in the period was monitored and their quality was validated. In particular, for each of them a reviewer was appointed and a BSCW enabled workflow was designed to automatically monitor its production phases.

Additionally, the Project Director initiated and maintained interactions with other 6th and 7th Framework projects and other R&D national/international programmes. Several new

collaborations with FARM related projects and organisations were established during the First D4Science World User Meeting. This meeting was carefully prepared to maximise its impact. The programme, the list of participants and the distributed project documentation were largely discussed in several phone conferences.
Two External Advisory Board Meetings were organised where the plan and achievements of the project were discussed with international experts. Quarterly Execution Board and Technical face-to-face meetings and a large number of weekly phone conferences were held to coordinate activities, examine developed solutions, and plan new actions.
From an administrative and financial management standpoint, the project coordinator ERCIM packaged and submitted the deliverables, contributed to the organization of plenary meetings and reviews, conducted the periodic reporting and the cost declarations under the NEF tool, distributed EC payments to partners and acted as the main point of contact between the consortium and the European Commission.
From a more focused effort and budget point of view, the consortium produced a global effort 22% higher than initially expected in terms of person-months and submitted a global financial claim 6% higher than initially expected for the full duration of the project. Half of the planned resources (49% exactly) were consumed in the first year, and it turned out that the second year required a significantly higher effort from the consortium in other to reach its set objectives. From a general standpoint, it can be concluded that D4Science delivered the expected set of quality results on time and on budget.

More information about the project can be found in the project Website: http://www.d4science.eu

2. Use and dissemination of foreground

2.1. Section A - Dissemination measures

The Communication and Dissemination activity evolved with the life of the project taking into account the different developments steps, achievements and user communities interested in the project outcomes. This evolutionary strategy will be preserved in the future taking into account the new contextual conditions. In particular, the D4Science Website will continue being operational by reporting the D4Science achievements and documentation and by serving its successor, D4Science-II. All the improvements on the web site, as implementing the site sections in order to achieve a more comprehensible structure, with a more accessible language, will also be applied to the D4Science materials and information.
The D4Science-II project will retain the same image (i.e. the logo and graphics) in order clearly show the continuity with its ancestor. Established collaborations with on-going projects, organizations and, more in general, with potential user communities will be maintained and strengthen.
Finally, the achievements of the D4Science project will continue to be disseminated through contributions to papers and conferences planned for the following months. The main relevant topics addressed/that will be addressed are:

- **e-Infrastructure Enabling Technology**

The gCube system is a Software Platform for the implementation and servicing of Virtual Research Environments (VREs). gCube is based on the Service Oriented Paradigm and a number of specifications and technologies that relate to the technological and functional areas addressed by its core services.
gCube provides not only a complete set of services to support the instantiation of VREs but also a set of sub-frameworks for the adaptation and expansion of its services and capabilities, so that the provided features can match the requirements of particular scientific application domains. Many are the innovative solutions introduced to support gCube's functionality, e.g. the resource model, the SOA enabling Information System, the On-Demand VRE generation, the Virtual Organization Manager, the Process Manager.

- **Information Management Services**
  gCube provides a set of services and frameworks for distributed organisation and retrieval of structured and unstructured content, which support the entire knowledge lifecycle within a Virtual Research Environment. The gCube Information Space is built on a unified, generic, flexible, modular information object model and is enabled by storage-management services that can exploit a diverse set of storage resources for placing payload, such as the Grid or local resources. It is fed by a powerful, flexible Import service that allows harvesting literally any kind of data and metadata. gCube has powerful multiple index forms (full text, geo-temporal, forward, vector, etc) supporting full incremental index updates and Distributed index creation and lookup. The gCube Search Engine performs full dynamic optimal allocation of Information Retrieval resources supporting at the same time custom operations on data and assuming no restriction in metadata content manifestations.

- **Virtual Research Environments**
  Virtual Research Environments lay at the core of scientific digital innovation. They provide collaborative frameworks enabling scientists to produce and exchange results with peers around the globe in cost-efficient manner. They not only offer capabilities for accessing cross-disciplinary data and knowledge but also provide services for exploiting a multiplicity of other disparate tools and computing resources enabling innovative analysis, simulation and domain specific knowledge generation processes. Realising operational VREs able to satisfy the needs of the very diverse scientific user communities imposes a paradigm shift in the way in which research organisations operate nowadays. This change involves at least three major areas: (i) the technology, (ii) the organizational model of the resources, and (iii) the human processes. These three areas are not independent, but any choice in one of them strongly influences and constrains the others. So progresses in one of these areas, e.g. the technological one as addressed by D4Science, are expected to stimulate modifications and adaptations in the others, and also to start possible further changes in that area itself.

- **Fisheries and Aquaculture Resource Management / Environmental Monitoring added value services**
  The availability of cross-domain resources enables creating new type of services for the integrated and simultaneous comparison, analysis, and processing of information. Several are the specialised services that can be introduced or enhanced to satisfy the more and more demanding requirements of communities. Examples, already implemented in D4Science, are new approaches to the generation of species distribution maps, new working environments for the production of compound reports, tools for time series integrated reallocation and curation.

- **e-Infrastructure Management**
  The cost of managing a production e-infrastructure strongly depends on the appropriateness of the defined policies and procedures and on the tools that are provided to support its management and monitoring. D4Science is the first example of infrastructure that simultaneously supports the notions of Virtual Organizations, Virtual Research Environments and controlled sharing of different type of resources, e.g. computing, service and data resources. The choices done, the services developed and the experiences collected are important elements for those that intend to develop and maintain either a generic infrastructure, like D4Science, or a domain specific one.

2.2. Section B - Exploitable foreground and plans for exploitation

The main tangible outcomes of the D4Science project are: the gCube software framework, the D4Science infrastructure which has been built by deploying this framework, the specific VREs set up to serve specific needs expressed by the user community, and the know-how gained on infrastructure development and operation. Each of these products has a potential market that range from large international organizations willing to create their own infrastructure, to groups of scientists willing to cooperate by sharing their resources, resource providers willing to make their resources more visible and widely exploitable, and other infrastructure that may be willing to exploit programmatically the resources and capability provided by the D4Science one.

The Consortium has started to analyse how to serve the variety of requested received from user communities in different application context for exploiting D4Science results. Three main options were identified for supporting the requests received:

a) creating a new infrastructure, a new VO and install own gCube Hosting Nodes(gHNs);
b) creating a new VO by joining the D4Science existing infrastructure and install own gHNs;
c) installing own gHNs and joining an existing VO in the D4Science Infrastructure.

Solution a) is the most expensive one since it requires that the new infrastructure is set up by installing appropriate services and then is monitored and managed. Solution b) is the easiest one. It requires a limited number of gHNs and the effort is limited to the management of the users of the VO. All the other services and tools are provided and maintained by the infrastructure. Solution c) is for free in terms of effort and system administration activities since it does not require the installation of any gCube Enabling Services. However, a political agreement is needed to join an existing VO. In this case, a request has to be sent to one of the D4Science VO administrators.

In trying to built an appropriate exploitation plan, a more in-depth exploitation analysis on the exploitation opportunities has been conducted for the user communities involved in the project activities. These communities have experimented the possibilities offered by D4Science to build collaborative environments. Both FARM and EM communities have additionally invested from their own resources to build synergies within their community, advocate the possibilities offered by D4Science, and deliver products that they can now already use. These UC experiences were presented at the D4Science World User Meeting, and the participants were specifically asked what products from D4Science would be of their interest, if they were available. A short inventory on participants' interest, but also their willingness to support and contribute to an e-Infrastructure revealed considerable interesting opportunities for data sharing, transformation and cross-domain integration.

Although it has not be possible until now to go through a detailed market survey (e.g. with the current mix of partners it is difficult to think of commercial offerings), the WUM inventory also revealed interesting opportunities for an e-Infrastructure. The potential market for D4Science, as emerged from this meeting, can be described as offering data, data-tools, services, and a platform:

Data storage and retrieval for clients:

- On-line document repositories;
- On-line data store for statistical, EO, environmental and biological data;
- Search and data-warehouse products.

Tools for data manipulation and collaborative processing by clients:

- Storage, processing and management of data, similar to an on-line database;
- Reporting tool, combining data;
- Shared reference data management;
- Geospatial analysis of data;
- Transformation of data between domains; e.g. transform SDMX data to a geospatial display format.

Services offered to clients:

- Consultancy services as for storage, processing and management of data, for private and public entities, offering on-demand quality information;
- Consultancy services to develop and manage VREs as a tool or data store.

Software platform for clients:

- The potential UC already has considerable experience in developing web-based software, web-services and databases. The WUM made clear that re-usability of artifacts or entire applications on the e-Infrastructure is high on the list of requirements, as this would offer a very cost-effective development model.
- Clients can develop software on a share agreement with the consortium;
- Clients can use the software to develop their own e-Infrastructure ;
- Offer interoperability solutions between e-Infrastructures.

Overwhelmingly, participants mentioned services for data management of EO data, biological and environmental information, and statistics as their main interest. There may have been a bias here towards data-centric prospects, as these participants formed a substantial group at the WUM. They are also generally more involved in data management, and thus can identify their needs better than other potential clients.

The participants also recommended that D4Science should focus on the one hand to offer services that help them solve their data-management issues, and at the other hand allow them to keep using their trusted environments and tools. These seemingly contrasting statements may be reconciled by further developing the interoperability capabilities of the D4Science infrastructure.

3. Conclusions

The main results of the D4Science project are:

- the gCube software framework;
- the D4Science infrastructure that has been built by deploying this framework;
- the VREs set up to serve specific needs expressed by the user communities;
- the know-how gained on infrastructure development and operation.

Each of these products has a potential market that ranges from large international organizations willing to create their own infrastructure, to groups of scientists interested in cooperating through sharing their resources, resource providers willing to make their resources more visible and widely exploitable, and other infrastructures that may plan to exploit programmatically the resources and capability provided by the D4Science one.

The above products enable a number of possible alternative solutions for serving collaboration and sharing requests coming from user communities, both in and out the EM and FARM domains. In particular, three main alternatives can be offered:

1. creating a new infrastructure, a new VO and install own gCube Hosting Nodes (gHNs);
2. creating a new VO by joining the D4Science existing infrastructure and install own gHNs;
3. installing own gHNs and joining an existing VO in the D4Science Infrastructure.

Solution 1 is the most expensive one since it requires that the new infrastructure be set up by installing appropriate services and then monitored and managed. Solution 2 is the easiest one. It requires a limited number of gHNs and an effort limited to the management of the users of the VO. All the other services and tools are provided and maintained by the infrastructure. Solution 3 is for free in terms of effort and system administration activities since it does not require installing gCube Enabling Services. However, a political agreement is needed to join an existing VO. In this case, a request has to be sent to one of the D4Science VO administrators.

Until now the D4Science infrastructure has been mainly exploited to serve needs of the EM and FARM communities. The experience done, and the feedback received by these and other related communities, e.g. the biodiversity one, have shown a strong willingness to support and contribute to an e-Infrastructure and have revealed considerable interesting opportunities for data sharing, transformation and cross-domain integration. The analysis of the feedback collected at the D4Science User Meeting, for example, highlighted that the main interesting aspects of the e-Infrastructure for these communities are: (i) data storage and retrieval of heterogeneous and multi-type content (e.g. documents; statistical, EO, environmental and biological data; earth observation products), (ii) tools for data manipulation and collaborative processing by clients (e.g. reporting tools, combining data; shared reference data management, geospatial analysis of data; transformation of data between domains); (iii) software platforms for clients which enable the re-usability of artifacts or entire applications in the context of an e-Infrastructure so to offer a very cost-effective development model.

Another benefit of the D4Science approach which emerged clearly in the course of the experimentation performed by the user communities is its capability to act as a catalyst for increasing the resource sharing, cooperation and cross-fertilisation among multiple communities. This feature is particularly important for all those domains where scientific communities are dealing with grand scientific challenges.

Despite the fact that until now the D4Science e-Infrastructure has mainly been experimented in the EO and FARM areas, it can easily support other domain areas. Recently, an experimentation of VREs supporting Culture Heritage communities has been initiated. The first results of this experimentation are very encouraging. A collaboration plan with other research groups external to the D4Science consortium is being investigated for extending gCube so to further address specific needs of these communities.

It is fair to conclude that D4Science has been a very active project during the two years of its EC-funded period and has produced many concrete results with an enthusiastic uptake from its end-user communities, all this on time and on budget. To continue this successful initiative, the European Commission has decided to put its trust in a very similar consortium that has started a follow-on project since October 2009, D4Science-II.