



PROJECT FINAL REPORT

Grant Agreement number: 239019
Project acronym: D4SCIENCE-II
Project title: *Data Infrastructures Ecosystem for Science*
Funding Scheme: CCPCSA
Period covered: from 01/10/2009 to 30/09/2011

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

Jessica MICHEL, project coordinator, ERCIM

Tel: +33 4 92 38 50 10

Fax: +33 4 92 38 50 11

E-mail: jessica.michel@ercim.eu

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

Project deliverable: DNA1.7

TABLE of CONTENTS

1. Final publishable summary report.....	4
2. Use and dissemination of foreground.....	17
2.1. Section A - Dissemination measures.....	17
2.2. Section B - Exploitable foreground and plans for exploitation.....	19
3. Conclusions	21
Appendix A - Dissemination (public)	22
Appendix B – Patents and Exploitable Foreground (confidential)	38
Appendix C - Report on societal implications	46

TABLE of FIGURES

Figure 1: The D4Science Ecosystem	4
Figure 2: gCube Website.....	8
Figure 3: D4Science gateway.....	10
Figure 4: Maps Generation in the AquaMaps portal.....	12
Figure 5: Vessel Transmitted Information VRE	13
Figure 6: D4Science Website	15

1. Final publishable summary report

Project Objectives

The major project aim was to develop technologies for enabling interoperation of diverse data e-Infrastructures that are running autonomously thereby creating an e-Infrastructure Ecosystem. This Ecosystem has the power to serve a significantly expanded set of communities dealing with multidisciplinary challenges whose solution is currently beyond reach. In particular, the project planned to create an initial ecosystem composed of at least the GENESI-DR and DRIVER repository e-Infrastructures, and other important thematic repositories maintained by international organizations, like INSPIRE, AquaMaps and FAO Infrastructure (The D4Science Ecosystem, Figure 1). By exploiting the resulting ecosystem, the project also aimed at supporting a number of Virtual Research Environments offering innovative functionality for facilitating scientific activities.

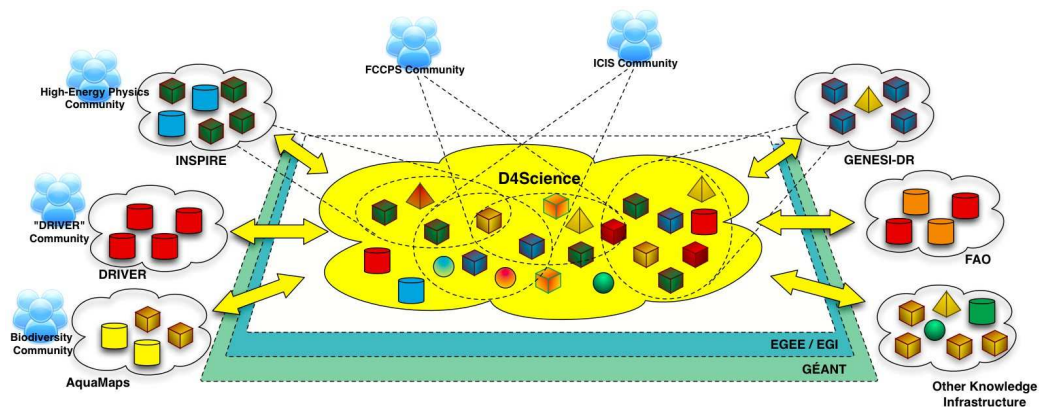


Figure 1: The D4Science Ecosystem

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 D4Science project management is jointly carried out by two individuals:

- Jessica Michel Assoumou – Administrative and Financial Director, 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 9238 5089 e-Mail: Jessica.Michel@ercim.eu

- Donatella Castelli – Project Director
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

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



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. gCube Development and Maintenance Activities
2. e-Infrastructure Management
3. Virtual Research Environments Development and Enrichment Activities
4. Communication, Dissemination and Training Activities
5. Coordination and Management Activities

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

i. gCube Development and Maintenance Activities

One of the major objectives of the project was to enhance the gCube¹ technological capabilities in order to enable D4Science to hold together and interoperate with all data e-Infrastructures that participate in the ecosystem. In particular, the project aimed at extending gCube with new mechanisms that not only enable interoperability but also minimize the need to use ad-hoc solutions for discovering, accessing and using resources published by various, heterogeneous data e-Infrastructures.

Main achievements

The gCube system was progressively enriched in the course of the project through a continuous delivery process which consisted of a total of 13 releases, i.e. 5 major releases and 8 maintenance releases.²

The production of these releases followed a strict cycle coordinated with the Service and Joint Research Activities. Software Integration, Testing, Distribution and Documentation tasks were performed in order to, respectively, take care of building and packaging the gCube software, check its deployment capabilities, functionalities and performances, and diffuse suitable documentation of the gCube software through a public web distribution site (<http://www2.gcube-system.org/>, Figure 2).

As a general policy driving the design and development, generic solutions to harmonization, integration and interoperability problems were preferred to specific ones when possible. *General-purpose frameworks* were integrated in the gCube system to address these problems. These frameworks facilitate reuse in multiple contexts thus enlarging the application domains served and reducing future development costs.

In particular, a number of *standards* were implemented and a set of *mediator frameworks* that can be configured to address different interoperability scenarios were developed. As a result of this activity the D4Science infrastructure is now able to interact with all those resource providers and consumers that implement the supported standards or can be served by the supported class of mediators.

¹ gCube is the D4Science e-infrastructure enabling system. Its development started in the D4Science project.

² The process of producing software releases and exploiting them to maintain and operate the infrastructure continues after the official end of the project.

In the data access area, the identified interoperability challenges led to design a completely new approach based on what was called *Open Content Management Architecture* (OCMA). This architecture set the road for the evolution of the D4Science Content Management services. Starting from the assumptions that (i) content is created, accessed, and distributed in units called documents, (ii) documents are grouped in collections, and (iii) collections are hosted in local management systems called repositories, OCMA defines the way for embracing and hiding the heterogeneity of the location where the content resides and how it is described and accessed. Such architecture also identifies classes of services and assigns to them a specific role. Along this view, a new suite of gCube components was implemented. Plugins were delivered for interfacing known services, e.g.: OAI-PMH compliant repositories.

A generic *gCube Data Transformation Service* (gDTS), responsible for transforming content and metadata among different formats and specifications, was also developed. gDTS lies on top of Content and Metadata Management services. It interoperates with these components in order to retrieve information objects and store the transformed ones. Transformations can be performed offline and on demand on a single object or on a group of objects. The version of gDTS released at the end of the project is also able to exploit the PE2ng, i.e. the Process Execution Engine, as a worker node of its operation, thus facilitating the composition of several transformations into more complex ones.

In the data discovery area, the existing rich set of Information Retrieval (IR) capabilities of the gCube platform was enriched and expanded in the direction of extending the interoperability of gCube Information Retrieval components with external infrastructures and platforms. Needs emerged where external data providers and search engines needed to be integrated in the flow of an IR operation. The direction taken in most of these situations was the creation of a mediator-based approach to identify instances of the external engine and to profile the external engine in the same way internal gCube search providers are profiled. According to this approach, an OpenSearch service, able to interoperate with this protocol was implemented.

In the process execution area, a Workflow and Execution Engine (PE2ng) was designed and implemented to serve the cause of bridging the gCube enabled nodes with platforms such EGI infrastructure, Condor, and Hadoop. Different cloud platforms were also interfaced in order to profit from elastic computing. This work was partially performed by exploiting synergies with the VENUS-C project. This extension represents a new big step towards the openness of the system.

In the security area, the authorization framework was extended with an Authorization service, an XACML Policy Repository and an XACML Callout Implementation. In particular, the adoption of XACML 2.0, which is an open standard, grants interoperability with external compatible platforms through exchange of policies with the respective XACML Policy Repositories.

Special attention was also dedicated to the scalability of solutions for the existing technology. In particular, the original gCube Information System was revised by making it able to support new standard solutions (WS-DAIX), the Information Retrieval components and their interaction model were revisited to allow better utilization of existing resources and the deployment model was modified by exploiting a new notion: the *virtual platform*. This is a model to be extended for transparently interfacing a potentially unlimited number of hosting environments. The gCube enabling technology was empowered with capabilities for dynamically instantiating platforms (along with their resources) compliant with such a model.

A considerable activity was also dedicated to implement appropriate *Presentation Services* offering the level of user interfaces usability required by the target user communities.

In the security area, an *Authorization framework*, composed by a pluggable Security Library, an XACML Policy Repository and a set of plug-ins for basic secure behaviours was released. In order to enable interoperability with external infrastructures, two extension packages were also implemented to perform controls based on the identity of incoming messages, and to provide VOMS credentials based on user, host or robot certificates.

A number of tools, e.g. BSCW and traditional tools, such as mailing lists and Wikis, TRAC and Subversion (SVN) were maintained for supporting a common environment that could facilitate interaction among the different teams and project tasks. A number of procedures and metrics were also elaborated for measuring the quality of development activities as a further mean to ensure the expected production quality level.

The screenshot displays the gCube Framework website. At the top, the logo 'gCUBE Framework' is visible. Below it, a banner features the text: 'gCube / Framework. gCube is a framework dedicated to scientists. It enables the declarative and interactive creation of transient Virtual Research Environments that aggregate and deploy on-demand content resources and application services by exploiting computational and storage resources of a grid infrastructure.' To the right of the text is an illustration of five stylized human figures holding hands. Below the banner is a navigation menu with links: Home, Success Stories, gCube Technologies, Documentation, Download, and About us. To the right of the menu is a login form with fields for Username and Password, and a Login button. Below the login form is a link: 'Forgot your password? Create an account'. The main content area is titled 'Overview' and contains two sections: 'gCube: the enabling e-Infrastructure framework view' and 'gCube: the data and process enabling framework view'. The right sidebar contains a 'NEWS and EVENTS' section with several release announcements for gCube versions 2.7.1, 2.7.0, 2.5.1, 2.5.0, and 2.4.1. At the bottom of the sidebar is a 'View All News' link and an announcement for the 'EGI User Forum 2011 Vilnius, Lithuania, 11-14 April 2011' with the EGI logo.

Figure 2: gCube Website

ii. e-Infrastructure Management

The objective of this area was to deploy and ensure the correct, continuous and effective operation of the resources shared through the D4Science Ecosystem while enabling their exploitation by the user communities. This includes the definition of procedures regulating the operation and enrichment of the infrastructure, the provision of appropriate tools for its monitoring and management, the infrastructure management and upgraded by periodically deploying new community-specific resources as well as more consolidated and extended releases of the supporting system, i.e., gCube. This area also includes integration, building testing, and certification of new major gCube releases.

Main achievements

The D4Science Infrastructure was deployed at the beginning of the project starting from the outcomes of the D4Science project. During the project lifetime it was gradually expanded by registering several additional resources, making it interoperable with several infrastructures, creating new Virtual Organizations (VOs) and new Virtual Research Environments (VREs). As a result of this activity, at the end of the project the infrastructure hosts *five Virtual Organizations* and *thirteen Virtual Research Environments*.

The management of the D4Science e-Infrastructure was facilitated by the definition and implementation of clear procedures for monitoring, accounting and incident management and by the formation of a specific support team.

A portal was developed to simplify the deployment of new Virtual Organizations (VOs) and VREs and a number of monitoring tools were deployed allowing administrator with different infrastructure roles to visualize the status of their resources and to be actively notified when problems occurred. In particular, a new tool for the *Service Availability Monitoring* was developed, which ensures the availability of the base gCube services functionalities, e.g. the Information System and the Search subsystem.

An accounting tool was also put in production providing relevant statistics about the users' exploitation of the infrastructure.

In order to reduce the impact of upgrades on the production environment, a *pre-production environment* was set up during the second year to validate the new developments and upgrades coming from the integration, testing and distribution team.

In order to increase the quality of service, it was decided to maintain two instances of the web portal acting as a point of access to the D4Science-II infrastructure: one offering the service to the User Community and one (called "*newportal*"³, Figure 3) dedicated to host the next version of the services including the new technologies released by the technical work packages. This strategy helped in reducing the downtime experienced by the user communities.

Suitable documentation of the gCube software through a public web distribution site was produced and continuously updated.

³ <https://newportal.d4science.research-infrastructures.eu/>

D4SCIENCE INFRASTRUCTURE Data Infrastructure Ecosystem for Science

Data e-Infrastructure gateway gCube 2.7.2 about gCube

Sign In

Screen Name

Password

Remember Me

[Create Account](#) [Forgot Password](#)

Welcome to the D4Science Data e-Infrastructure gateway

This gateway is an access point to a number of **gCube applications** operated by the D4Science e-infrastructure. These applications offer specialized functionality for the management, processing, and visualization of scientific data and textual content. They exploit the D4science capability to interoperate with heterogeneous e-infrastructures ranging from traditional data archives and institutional repositories to more advanced grid and private/public cloud resources providers.

Among the e-infrastructures bridged by the D4Science e-Infrastructure, the following deserve a special mention: the **European Grid Infrastructure (EGI)**, the **Earth Science data** and processing commodities (**GENESI-DEC**), the **Cloud Infrastructure for Science (Venus-C)**, the **European Digital Repositories Infrastructure (Driver)**, the **High Energy Physics Gateway (Inspire)**, the **Marine Species Distribution Infrastructure (AquaMaps)**, and the **FAO GeoNetwork**.

gCube applications are made available through **Virtual Research Environments (VREs)**. Virtual Research Environments are collaborative environments enabling scientists and practitioners 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 their users with a rich array of services enabling innovative analysis, visualization and domain specific knowledge generation processes.

You have to **be registered** in order to use the facilities offered by the available VREs by following the simple policy reported below:

- If you are a member of any of the existing communities, you can request access to the community's VREs;
- If you are not a member of any of the supported communities, you can request access to one or more gCube Apps VREs. The gCube Apps VREs offer a **free-to-use** environment and an **unlimited** use of **storage** and **computational capabilities**. The data you upload and generate are kept private to your workspace until you decide to share with other members of any gCube Apps VRE.

The current communities using the e-Infrastructure are:

- **FARM**, the *Fisheries and Aquaculture Resource Management community*, bringing together the "Food and Agriculture Organization of the United Nations" (FAO) - Italy, the "International Center for Living Aquatic Resources Management" ([WorldFish Center](#)) - Malaysia and "Fishbase Information & Research Group Inc" (FIN) - The Philippines.
- **DRIVER**, the *community served by the DRIVER infrastructure*, to enhance the on-line services and the off-line capabilities the DRIVER portal, currently aggregating approximately more than 5 million documents from 300+ repositories in 40+ countries.
- **INSPIRE**, the *HEP community served by the INSPIRE infrastructure*, to extend the INSPIRE High Energy Physics repository with advanced search and retrieval functionality based on computationally intensive processing of bibliometric information.

If you want to host your community in the D4Science e-Infrastructure, if you want to become either a service provider or a data provider, if you want to exploit programmatically the resources aggregated by the D4Science e-Infrastructure you can [email](#) us.

You can read more about D4Science on the website (<http://www.d4science.eu>) and learn more about the project, its objectives and the people behind it.

D4Science is partially funded by the European Commission under its Seventh Framework Programme.

Home | D4Science 2008-2011 | Privacy

Figure 3: D4Science gateway

iii. Virtual Research Environments Development and Enrichment Activities

This area aimed at realising and operating Virtual Research Environments serving very different application scenarios. These scenarios exploit resources that are provided by the different e-Infrastructures participating in the ecosystem. Some of these VREs were expected to be consumed through end-user interfaces, while others were expected to be consumed programmatically by third-party services through APIs.

Main achievements

These VREs played a key role in the project since they enabled to validate the effectiveness of the support offered by the infrastructure and to convey to the community concrete examples of possible innovative exploitations.

In particular:

- *INSPIRE VRE*: This VRE was created to offer to the INSPIRE High Energy Physics Information System, through APIs, a set of specific functionalities which require large scale processing capabilities. The development around this VRE focused on three algorithmically stable areas: *metadata extraction*, *OCR-ing* and *full text indexing*. Despite solutions were addressing requirement related to the INSPIRE system, the project aimed at developing more general components that could also have been reused to serve the needs of other infrastructure customers. After an experimentation and pilot development performed in the first year extensive developments and tests were performed in order to achieve high scalability. JDL and grid resources were used for OCR and metadata extraction. In particular, full text indexing was performed by exploiting the Hadoop based parallel processing. Tests with the Hadoop cluster allowed parallel processing of up to 23 thousand documents in 2h time.
- *DRIVER VRE*: This VRE was set up to provide functionality to the DRIVER and other D-Net⁴ empowered infrastructures. It offers back-end facilities to this infrastructures by making available D4Science accessible content through the OAI-PMH standard protocol and by providing typical library functionality that requires large processing capabilities. Examples of this functionality are *thumbnailing* with custom thumbnailers and extraction of text from PDF documents maintained in distributed repositories through *OCR-ing* so to enable multidimensional search capabilities. The latter facility required the implementation of a Resource Discovery component to search for documents through the web (following the link indicated in the aggregated metadata records), starting from a given set of resources up to a given search depth. The PDF documents so retrieved can now be processed by applying the OCR services developed in the context of the INSPIRE scenario to provide the necessary information to the D-Net indexing service. A series of experiments for the tuning of this facility were performed considering a limited number of DRIVER collections. These experiments highlighted the need for greater exploitation of the D4Science computational back-end and for access to context-free processing over grid and cloud computing, to be offered through the Process Execution Engine of gCube.
- *AquaMaps VRE*: This VRE offers a comprehensive working environment supporting the production of *species distribution maps* via predictive algorithms matching species characteristics with environmental data (Figure 4). It was deployed in the production infrastructure at the beginning of the project and since then it was constantly updated by acting on both user-oriented and systemic functionalities. Examples of the former are the integration of a richer set of algorithms for computing the species distributions, the provision of capabilities for selecting the version of the dataset to be processed in order to produce the distribution maps, and the implementation of a *Transect chart framework*, which includes tools for charts visualization and labelling usable to analyse the species distributions over a special area. Examples of the latter are the integration of a number of processing platforms, including Cloud ones, for improving the performance of calculations, the implementation of an HTTP-based interface for programmatically exposing the AquaMaps objects produced through the suite, and the production of GIS maps that can be disseminated through *GeoServer*. Because of the effectiveness of the release environment, the AquaMaps community raised pretty soon the need to sync the gCube AquaMaps environment with the official AquaMaps service (www.aquamaps.org).
- *FCCPS VRE*: The Fishery Country Profiles Production System VRE is focused on the production of country reports in the context of fisheries management. This VRE has been operated since the beginning of the D4Science production infrastructure. In the course of the project it has been improved to better support the collaborative activities. A *Business Document Workflow Suite*, i.e., a comprehensive working environment supporting the

⁴ D-Net is the DRIVER supporting system

definition and execution of workflows governing the collaborative production of rich documents, was finalised. Other activities were performed to better meet and integrate the on-going practices of the community working at the production of the Country Reports. Examples of such activities are the customization of the D4Science report model to accommodate the features needed to support the Fishery Country Sector branch of the FiMES schema, and the enhancement of the Report and Template Portlets to create an online version of the Fishery and Aquaculture Country Profile Word template and make the export possible by exploiting the FAO converter tool.

- **ICIS VRE:** This VRE supports fisheries statisticians by providing them with a set of tools to *curate and manage catch statistics and related data*. It has been operated since the beginning of the D4Science production infrastructure. Pretty soon in the course of the project it emerged that the application operated by this VRE might benefit from the bunch of facilities offered by the *SDMX* (Statistical Data and Metadata eXchange) standard, e.g. for code list linking and time series publishing. Requirements on this innovative feature were analysed and web services supporting its provision were developed. Moreover, the capabilities of this VRE were largely expanded by integrating the existing application with *R* and with other VREs, such as *VTI*, and by making it interoperable with *FLOD* and *GeoServer*. It was also enhanced with graphing facilities and with the provision of a more friendly User Interface.

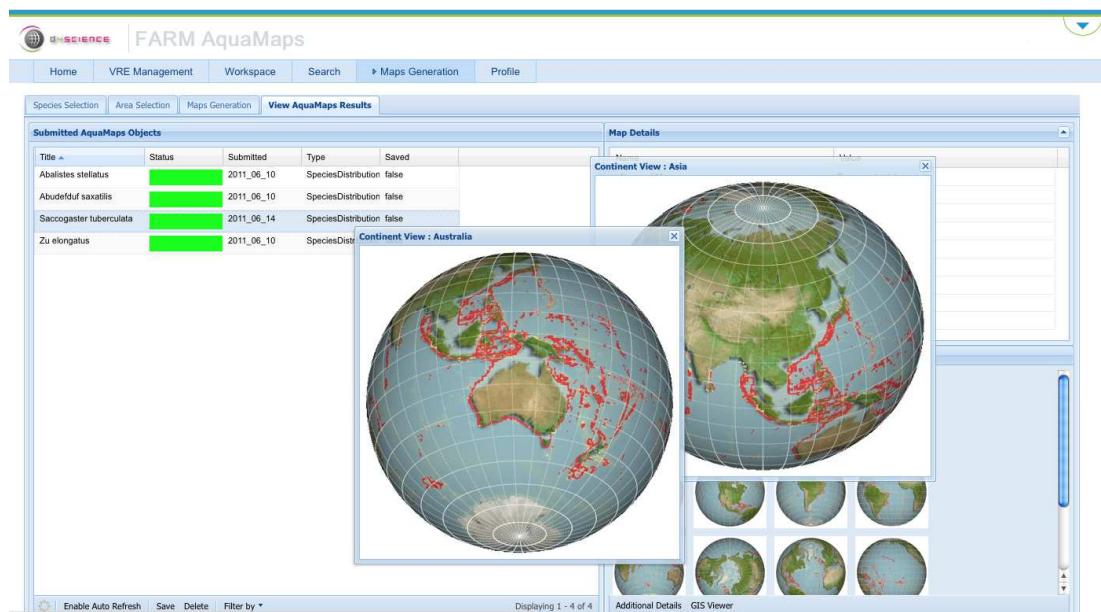


Figure 4: Maps Generation in the AquaMaps portal

The more in depth analysis of the five application scenarios, the availability of new infrastructural resources and the development of new capabilities, stimulated a number of additional activities originally unplanned. Major results produced by these activities are:

- **VTI VRE:** this VRE, named Vessel Transmitted Information, is a spin-off of the ICIS VRE. It was created to specifically target the *management of Vessel Position Records* to extract monthly fishing effort per unit area (Figure 5). The scenario requires that statistical data can be merged with environmental data, e.g. bathymetry extracted from GENESI-DEC. This was proven feasible. The specific position of VTI, at the exact intersection between the spatial and statistical communities, offers real and valuable services to a potentially very large community. As part of the development related to this VRE, new software artefacts were implemented, e.g. software libraries crunching vessel records for guessing fishing activity.

- *OpenSDMX environment*: this environment, exploited in the ICIS VRE, aims to leverage the power of D4Science technologies for the international statistical data collection community. It addresses the provision of mechanisms facilitating the management of code lists and mechanisms aiming at reinforcing the management of time series objects by relying on the SDMX standard. In particular, it comprises an SDMX Publisher offering a RESTful interface serving FAO Codelists. It also supports an SDMX DSD's for the Fisheries and Aquaculture domains, currently used by e.g. FAO and Eurostat. Collaboration agreements at the level of software development are being drafted with other organisations to join forces on continuing its developments also after the end of the D4Science-II project.
- *FLOD: Fisheries Linked Open Data Knowledge Base* is a knowledge base dedicated to capture semantic knowledge on entities ranging from species to geospatial entities. At the end of the project it consists of more than 20 rdf linked datasets in the fisheries domain, each equipped with a set of services to maintain and consume the contents. FLOD data cross the spatial, statistics and marine legislation domains. The integration was achieved using semantic technologies.
- *SPREAD: 'Spatial REAllocation of Data'* is a demonstrator built by exploiting D4Science capabilities. It deals with the process of transforming data from one spatial resolution to another one. By relying on a number of facilities and repositories offering geospatial and other kind of data hosted in the Ecosystem, this demonstrator offers functionality for distinguishing catches in the High Seas from catches within the EEZs (one of the UN General Assembly recommendations). In the context of this demonstrator, a number of requirements related to quantity and quality of catch and related data to be managed to serve these scenarios were fulfilled.
- *TimeSeries suite*: This is a comprehensive working environment supporting the management and curation of time series objects. It was developed during the reporting period as part of the ICIS VREs. Among the others, it makes available facilities for dealing with errors in data during the curation phase, supporting the management of code lists, producing a rich array of graphs from time series; and transforming time series data to a set of GIS layers.

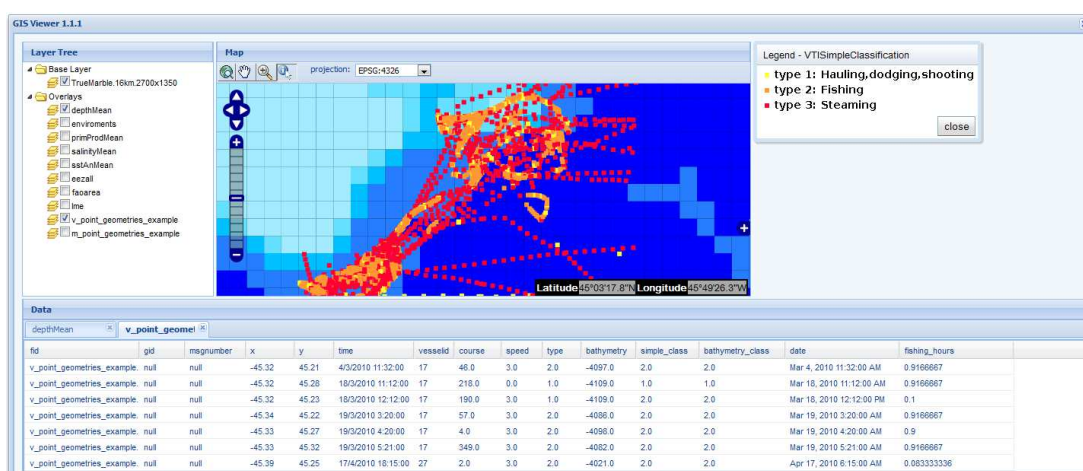


Figure 5: Vessel Transmitted Information VRE

iv. Communication, Dissemination and Training Activities

This area aimed at fostering the growth of a D4Science-II community by disseminating project results and knowledge to raise awareness within interested scientific circles. It also

was in charge of training software engineers, VO and VRE administrators and end-users in the configuration and exploitation of the gCube system and of the D4Science infrastructure facilities.

Main achievements

In order to raise awareness and disseminate project results among the largest pool of potential users and user communities, live D4Science project (Figure 6), and gCube web sites (Figure 2) were maintained to reflect the on-going work of the project and its interaction with scientific communities. Project partners were actively involved in communications at scientific events such as talks, demos, panels, discussions, presentations, tutorials, and poster sections. A range of presentations, brochures and articles in scientific literature were produced to convey the complex technical details of the project in a language that clearly showed the power and reach of the project. Three videos were also produced to effectively communicate relevant information and offer an interactive and engaging experience to the visitors of the D4Science-II website. Finally, scientific papers were published in scholarly, scientific and academic journals in order to increase visibility among research communities and target groups.

D4Science-II collaborated with different projects and R&D programmes striving for similar or complementary objectives. These collaborations contributed both to increase awareness and visibility of the D4Science-II project and to enlarge the planned ecosystem.

Relevant to the establishment of these collaboration was the organization of a number of events, like the D4Science World User Meeting, attended by more than 80 international fisheries and biodiversity delegates, a meeting with ESFRI Environmental projects, and another one, involving digital libraries and repository experts. Stimulated by recommendations received from the project's External Advisory Board, the Project Management Board discussed extensively the sustainability of the several D4Science-II products, e.g. D4Science infrastructure, gCube system, VREs, and various developed tools. A workshop dedicated to discuss this specific aspects with a larger community was organised towards the end of the project. Collaboration activities were also performed with the ERINA+ project, which tested its methodology for socio-economic impact assessment by analysing the exploitable results of the AquaMaps VRE.

As a further contribute to outreach, on-line tools and electronic material for support of events and off-line training were produced. Several training sessions to external and internal to the project audience were held and others were planned to be hold in near future also outside the project lifetime.

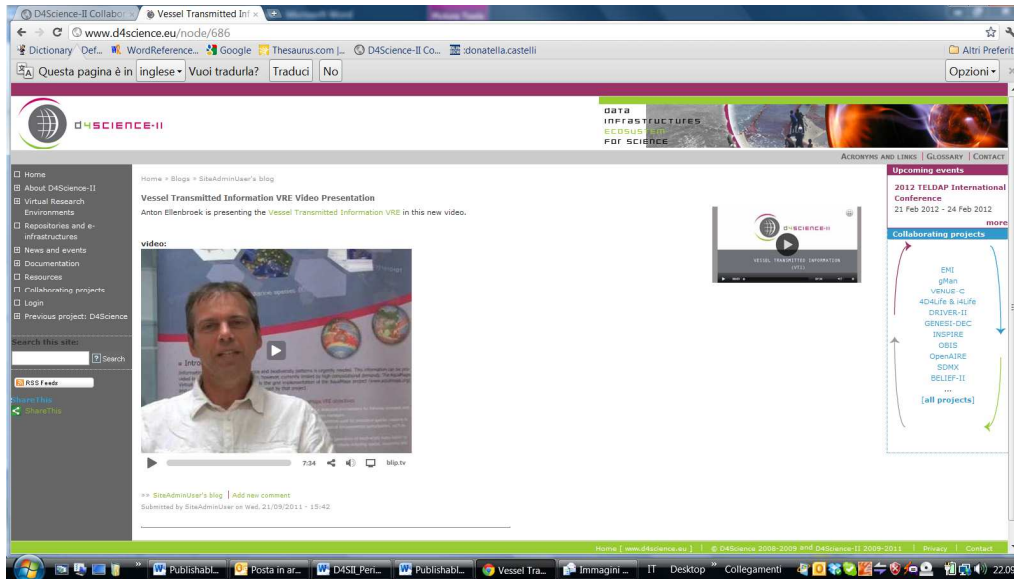


Figure 6: D4Science Website

v. Coordination and Management Activities

This last area was concerned with all the actions dedicated to the overall management of the project in its scientific, technological and administrative aspects. In particular, it included the definition of a detailed work plan, the monitoring and assessment of the progresses of the project with respect to this work plan, the measurement of the quality of the activities performed and the timeliness application of corrective actions. It also included careful monitoring of the resources and financial expenditures and the fulfillment of contractual obligations and reporting.

Main achievements

Activities were continuously supervised from the scientific, technical and financial points of view in order to assure appropriate quality and progresses. To help in this process, a Quality Assurance Task Force, appointed at the beginning of the project, defined a project quality plan and monitored it. The defined procedures concerned 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. These procedures were made available in a wiki website open to all the project partners that also contained an up-to-dated progress status of the project deliverables.

The Scientific and technical directors operated to achieve the completion of the activities according to the established plans. Particular attention was dedicated to assure the proper level of interaction between the technologists and the representatives of both other e-Infrastructures participating in the ecosystem and scientific communities exploiting the VREs. These and other project activities were coordinated through periodic physical meetings and phone conferences.

Many co-operations agreements with other EU projects and initiatives were established, often as a result of networking events organised by the project. For example, as an outcome of the meeting organized with the ESFRI Environmental projects and of the ESFRI ENV Thematic working group the D4Science infrastructure was selected as underlying platform in the context of two new 7th Framework projects: ENVRI and EUBrazilOpenBio. The former, which will start on November 2011, aims at developing mechanisms for the integration and harmonisation of resources from environmental infrastructures by supporting the definition of unifying views over heterogeneous information spaces. The latter, which will start on October 2011, aims at deploying an open-access platform supporting the needs and requirements of the biodiversity scientific community by federating and integrating existing European and Brazilian infrastructures and resources.

An intense activity was initiated towards the end of the project to identify solutions for the sustainability of the D4Science Infrastructure. In particular, a collaboration activity was established with a team of experts of the ERINA+ Project to develop a preliminary “socio-economic impact assessment”, through the analysis of one of the project’s exploitable results: the AquaMaps Virtual Research Environment. Moreover, a 2.5 day workshop dedicated to address the unique sustainability issues confronted by the infrastructure was organized in near Nice, France, on September 2011. The outcome of this workshop has been reported into the public deliverable “DNA1.5 *Action Plan for Exploitation and Sustainability.*”

2. Use and dissemination of foreground

2.1. Section A - Dissemination measures

In order to raise awareness and disseminate project results among the largest pool of potential users and user communities, the following dissemination measures were performed:

- Creation of a set of graphic materials:
One triptych brochure and six inserts with an A5 size, to be stored within it. The inserts address these topics: gCube, AquaMaps, ICIS, FCCPS, DRIVER, INSPIRE. These materials were prepared to be used on different external events where D4Science-II team members participated aiming to transmit a clear overall picture of the project.
- Following the same visual identity, three posters were created on the following topics:
 - i. Generic poster about D4Science-II, its objectives and components
 - ii. Poster on the ICIS VRE
 - iii. Poster on gCube PE2ng

These posters were used in different conferences and adapted to each event necessity.

- A high-level explanation of the D4Science-II project was developed for reaching a wide audience and encouraging understanding of the project by a large community, bearing in mind that people who come for the first time to the project and can be potential users of it, do not necessarily have an accurate knowledge about the more technical parts.
- A project website was created and maintained. During the whole project period, 90 blog entries were published in the D4Science-II blog. They reported on project activities, highlight events, presentations, deliverables, milestones and promoted website's novelties. With different metrics, the project members have been able to monitor and evaluate the effectiveness of the communication activities. Figures were extracted from Google Analytics, a service offered by Google that generates detailed statistics about the visitors to a website. These statistics show that the regularly updated D4Science-II block page was the most frequented. The site reflected the ongoing work of the project and its interaction with scientific communities.
- A gCube website, describing the system technology and reporting its advances, was also created and maintained operational. Thanks to the lively web presence the number of visitors was fairly stable during the full period with visits from 82 countries. At the time of preparation of this report (Oct. 2011) this site has recorded 2353 software component downloads.
- Project partners participation to events, listed in the dynamic web calendar, was reported in the form of blog stories with attachments to the presented material. There were 41 selected dissemination events where D4Science-II was present and actively involved in communications such as talks, demos, panels, discussions, presentations, tutorials, poster sections. A range of presentations, brochures and articles in scientific literature were produced to convey the complex technical details of the project in a language that clearly showed the power and reach of the project. All materials prepared for conferences and

meetings were accessible - depending on their nature - either on the D4Science-II project website or on the internal project CMS.

- Free WEB 2.0 tools, such as social media sites, were used to create a powerful media project presence. Blip TV and YouTube (as a secondary channel) were used to add D4Science-II project videos. 3 videos were filmed and edited:
 - D4Science-II video about GRDI 2020 meets the D4Science Project;
 - D4Science-II video about Fishery Country Profiles Production System;
 - D4Science-II video about Vessel Transmitted Information.

With these videos the project effectively communicated relevant and complex information to the visitors of the D4Science-II website. The videos offered a more interactive and engaging experience, keeping visitors on the site longer.

- A monthly news digest was also introduced. It summarized the news published on the D4Science-II blog and was sent to all project contacts collected in the course of the different D4Science-II events.
- A long-standing technique for generating an image of authority, excellence and professionalism in scientific circles is through the publishing of papers in scholarly, scientific and academic journals. During the full period of the project the D4Science-II team worked on publishing papers with the objective of increasing visibility among research communities and target groups. 14 scientific publications were disseminated and are available in print and online.

2.2. Section B - Exploitable foreground and plans for exploitation

At the end of the D4Science II project, two distinct scientific communities are involved in exploiting and testing the D4Science infrastructure:

- FARM collect and analyze data on fisheries, biodiversity and ocean observation;
- INSPIRE and DRIVER develop computationally demanding e-library functions such as processing documents and computing bibliometrics.

The current exploitation by these communities can be grouped into two major areas:

- Distributed Open Source software development, pooled into one single infrastructure, fostering the adoption of standards, with tools/systems such as OpenSDMX and FLOD;
- Development and usage of Virtual Research Environments, with the aim of building resources for Communities of Practice, developing convincing demonstrations for other potential users, and fostering wider engagement of the community, through products such as Aquamaps, FCPPS, ICIS for potential use with visualisation tools such as maps or dynamic charts (e.g. Gapminder) and analytical tools (e.g. R).

These exploitation areas would be developed to further address and resolve an interlinking set of issues for the science/user community, including overcoming data collection and sharing problems, better meeting the need to generate reliable harmonized quality structured data sets, opening access to closed data repositories, and overcoming the lack of computing power amongst researcher, data users and resource managers. The practical aspects of these aims can be further defined to meet specific demands by the science/user community, including:

- The development of code list management and dissemination services;
- Archival services;
- Production and access to a library of on-demand Oceans observation satellite images;
- The availability of harmonized, qualified, and timely shared Time series of key data sets;
- Improved cross-domain analytical capacities, including interactive mapping capacities and services, and streamlined flow among analytical tools such as R and Gapminder;
- Improved and more integrated modeling capacities, including the modeling of species distribution and biodiversity with flexible space and time resolution, of fishing activity from vessel-transmitted Information, and of reallocation of catch statistics across species, stocks, fleets, reporting zones and countries.

These would be developed according to fundamental principles of supporting open access and free use of platforms and systems (with options for various fees/licensing for specific added value services, and for use of system components outside the primary communities – see next section); ensuring access and use of shared data and tools fully meeting users' confidentiality and security requirements; and greatly enhancing access to greater computing power and capacity.

To enable this to be achieved, a longer-term plan for system sustainability is required. While recognising the key role that core funding for D4Science has played in establishing the working elements of the system, if these are to be applied widely and well, the system maintained and updated, features and services developed and improved, and users guaranteed a reliable and ‘future-proofed’ resource, a pragmatic perspective is required to identify and develop a more diverse strategy for funding and service development. This would recognise the diversity of potential products and services, addressing various audience types, including users of infrastructure assets, software developers, data managers, statisticians, scientists and policy makers.

3. Conclusions

The main results of the D4Science-II project are:

- the gCube software framework release 2.5.1;⁵
- the D4Science Ecosystem, operated through the gCube framework, interoperating with several other infrastructures;
- thirteen Virtual Research Environments supported by the D4Science Infrastructure offering services to users through dedicated portals and to other services through specific APIs;
- a number of specific tools for curation, management and processing of data that can be exploited in many different application contexts.

Each of these products has a potential market that ranges from large international organizations willing to create their own infrastructure as well as companies interested in either developing services on top of the infrastructure or offering other infrastructure services (e.g. Cloud services) to D4Science, to groups of scientists interested in innovative and more competitive environments where to perform their experiments and generate new and better quality knowledge.

The D4Science-II project has also largely contributed to produce a considerable know-how on infrastructure interoperability development. Generic interoperability solutions have been introduced whenever possible. Several issues have been addressed and validation activities have been performed. Also, interoperability issues that are not usually known in the literature have been highlighted through concrete development and experimentation activities

The D4Science products have also acted as catalysts for other communities/organizations that have shown interest in them and have decided to join with the D4Science partners in order to move forward the development of these products.

⁵ The activity of releasing new gCube versions continued after the project official end. At the time of finalising this deliverable (November 2011) gCube 2.7.0 was available.

Appendix A - Dissemination (public)

A1: LIST OF SCIENTIFIC (PEER REVIEWED) PUBLICATIONS, STARTING WITH THE MOST IMPORTANT ONES

NO.	Title	Main author	Title of the periodical or the series	Number, date or frequency	Publisher	Place of publication	Year of publication	Relevant pages	Permanent identifiers (if available)	Is/Will open access ⁶ provided to this publication?
1	D4Science-II - An e-Infrastructure Ecosystem for Science	Donatella Castelli, Marc Taconet and Virginie Viollier	ERCIM News 79 Special theme: Towards Green ICT	October 2009	ERCIM	Europe	2009	P.9	Link	Yes
2	Functional Adaptivity for Digital Library Services in e-Infrastructures: The gCube Approach in proceedings of the Research and Advanced Technology for Digital Libraries,	Simeoni F., Candela L., Lievens D., Pagano P. and Simi M.	13th European Conference, ECDL 2009	October 2, 2009		Europe	2009	P51-62	Link	Yes
3	On-demand virtual research environments and the changing roles	Candela L., Castelli D., Pagano P	Library Hi Tech, vol. 27 (2)	2009	Emerald Group Publishing	Europe	2009	P 239 - 251	Link	Yes

⁶ Open Access is defined as free of charge access for anyone via the internet. Please answer "yes" if the open access to the publication is already established and also if the embargo period for open access is not yet over but you intend to establish open access afterwards.

	of librarians				Limited					
4	Building an e-Infrastructure for Capture Statistics and Species Distribution Modeling: the D4Science Approach	Candela L., Castelli D., Ellenbroek A., Pagano P., and Taconet M.	International Conference on Marine Data and Information Systems (IMDIS 2010)	March 2010		Europe	2010		Link	Yes
5	Making Virtual Research Environments in the Cloud a Reality: the gCube Approach	Leonardo Candela, Donatella Castelli and Pasquale Pagano	ERCIM News 83 Special theme: Cloud Computing	October 2010	ERCIM	Europe	2010	P46	Link	Yes
6	A count in the dark	Pauly, D. and R. Froese	Nature Geoscience 3	2010		Europe	2010	P662-663	Link	Yes
7	Deploying general-purpose virtual research environments for humanities research	Blanke T., Candela L., Hedges M., Priddy M., Simeoni F	Philosophical Transactions of the Royal Society of London Series A- Mathematical Physical and Engineering Sciences, vol. 368	2010	Royal Society Publishing	Europe	2010	P 3813 - 3828	Link	Yes
8	The gCube interoperability framework. In: 2nd DL.org Workshop - Making Digital Libraries Interoperable	Candela L., Kakalettris G., Pagano P., Papanikos G., Simeoni F.	2nd DL.org Workshop - Making Digital Libraries Interoperable	9-10 September 2010		Europe	2010	Proceedings : P35 - 42		Yes
9	An event-centric	Castelli D.,	Digital	2010		Europe	2010	P79 -	Link	Yes

	provenance model for digital libraries	Candela L., Manghi P., Pagano P., Tang C., Thanos C.	Libraries - 6th Italian Research Conference. Communications in Computer and Information Science, vol. 91 - Part 3					88		
10	D4Science: An e-Infrastructure for Facilitating Fisheries and Aquaculture Resource Management	Donatella Castelli, Anton Ellenbroek, Pasquale Pagano and Marc Taconet.	Abstract, presented at 22nd International CODATA Conference	24-27 October, 2010		CapeTown, South Africa.	2010		Link	Yes
11	An Event-Centric Provenance Model for Digital Libraries	Castelli, Leonardo Candela, Paolo Manghi, Pasquale Pagano, Cristina Tang, Costantino Thanos	Digital Libraries - 6th Italian Research Conference	28-29 January 2010		Europe	2010	P 91, 79-88	Link	Yes
12	Building Scientific Workflows for the Fisheries and Aquaculture Management Community	Pagano P., Candela L., and Andrade P.	5th EGEE User Forum	April 2010		Europe	2010		Link	Yes
13	Potential contribution of the Internet to a global community of practice for	Serge M. Garcia	ICES Journal of Marine Science	January 2011		Europe	2011		Link	Yes

	fishery management, Internet to a global community of practice for fishery management.									
14	The D4Science Approach toward Grid Resource Sharing: The Species Occurrence Maps Generation Case.	Candela L., Pagano P.	Data Driven e-Science - Use Cases and Successful Applications of Distributed Computing Infrastructures (ISGC 2010)	January, 2011		Europe	2011	P. 225-238	Link	Yes

A2: LIST OF DISSEMINATION ACTIVITIES

NO.	Title	Date	Place
1	World User Meeting	25-26 November 2009	Rome, Italy
2	Italian Research Conference on Digital Libraries	28-29 January 2010	Padua, Italy
3	2 nd meeting of the Italian Working Group on Digital Cultural Heritage E-Infrastructure	01 March 2010	Pisa, Italy
4	International Conference on Marine Data and Information Systems (IMDIS)	29-31 March 2010	Paris, France
5	International Symposium on Grid Computing	5-12 March 2010	Taipei, Taiwan
6	6th European Conference on Research Infrastructures (ECRI2010)	23-24 March 2010	Barcelona, Spain
7	D4Science meets ESFRI & RI Projects: Workshop	17 Mar 2010	Brussels, Belgium
8	Second Meeting of the Italian Working Group on Digital Cultural Heritage e-Infrastructure	1st April, 2010	Pisa, Italy
9	HASTAC 2010: Grand Challenges and Global Innovations Conference	15-17 April, 2010	London, UK
10	5th EGEE User Forum,	12-16 April- 2010	Uppsala, Sweden
11	The 5th international conference on Open Repositories (OR2010)	6-9 July 2010	Madrid, Spain
12	DH2010	7-10 July 2010	London, UK
13	ICES Symposium on the Collection and Interpretation of Fishery Dependent Data	23-26 August 2010	Galway, Ireland
14	The 2nd DL.org workshop (ECDL2010)	9-10 September 2010	Glasgow, Scotland, UK
15	The FOSS4G conference for Open Source Geospatial Software	6– 9 September 2010	Barcelona, Spain
16	EGI Technical Forum 2010	4-17 September 2010	Amsterdam, The Netherlands
17	ICT 2010	27-29 September 2010	Brussels, Belgium
18	Workshop with the Environmental projects	30 September 2010	Brussels, Belgium
19	DL.org Autumn School on Digital Libraries and Digital Repositories: Modelling, Best Practices, and Interoperability	03-08 October 2010	Athens, Greece
20	D4Science Demonstration at Theory and Practice in Digital Libraries: A European Approach (DL.org Workshop)	13 October 2010	Athens, Greece
21	22nd CODATA International Conference, Scientific Data and Sustainable Development	24-27 October 2010	Cape Town, South Africa
22	Workshop on Theory and Practice in Digital Libraries: A European Approach	12-13 December 2010	Athens, Greece

23	Workshop on Digital Repositories – Linked Open Data – the possible Role of D4Science	16-17 December 2010	Rome, Italy
24	5th Session of the IODE Group of Experts on Biological and Chemical Data Management and Exchange Practices	17-20 January 2011	Oostende, Belgium
25	Information Systems Architecture Forum – Cloud Computing and FAO	25 January 2011	Rome, Italy
26	29th Session of Committee on Fisheries (COFI)	31 January – 4 February 2011	Rome, Italy
27	Third Meeting of Regional Fishery Body Secretariats Network (RSN 3)	07-08 February 2011	Rome, Italy
28	Metafor Project Workshop: Using the Metafor Common Information Model (CIM) to store, discover and locate climate modelling data	14 Mar 2011	Abingdon, United Kingdom
29	TELDAP International Conference	16-19 March 2011	Taipei, Taiwan
30	International Symposium on Grids and Clouds (ISGC 2011)	21-25 March 2011	Taipei, Taiwan
31	European Grid Infrastructure (EGI) User Forum	11-14 April 2011	Vilnius, Lithuania
32	Presentation of D4S results in the Erina Workshop at Egi forum	12 April 2011	Vilnius, Lithuania
33	SDMX Global Conference 2011	2-4 May 2011	Washington DC, United States
34	Presentation of ICIS, the Integrated Catch Information System Virtual Research Environment (VRE) and the D4Science Supporting Technology for FAO staff	17 May 2011	Rome, Italy
35	Extreme Large Database Workshop	9-10 June 2011	Edinburgh, United Kingdom
36	Presentation of gCube system evolution and its architecture	11 June 2011	Athens, Greece
37	Sustainability Workshop	12-14 September 2011	Cagnes-sur-Mer, France
38	EGI Technical Forum 2011	19-23 Sept 2011	Lyon, France
39	Presentation of the paper “An approach to Virtual Research Environment user interfaces dynamic construction” by Massimiliano Assante, Pasquale Pagano, Leonardo Candela, Federico De Faveri and Lucio Lelii TPD L 2011 Conference	26-29 Sept 2011	Berlin, Germany
40	Presentation “Taming Data Sharing and Re-using in the e-Infrastructures realm” by Pasquale Pagano at “Global Research Data Infrastructures: The Big Data Challenge”	18-19 Oct 2011	Brussels, Belgium
41	Presentation “e-Infrastructures for the Future Ocean” by Donatella Castelli, at “Workshop on Scientific Data Infrastructures in the European Parliament”	20 Oct 2011	Brussels, Belgium

Dissemination material

D4Science logo



General information

- [D4Science-II brochure](#)
- [AquaMaps Insert](#)
- [FCCPS Insert](#)
- [ICIS Insert](#)
- [Driver Insert](#)
- [Inspire Insert](#)
- [gCube Insert](#)

FLYERS: communications team has produced a D4Science-II flyer and five inserts during the M1-M12 period. As the project developed there was a need to rewrite and redesign these previously prepared materials. This revision has been undertaken during the M15-16 period. The flyer gives an overview of the project together with its objectives and context. The flyer and the inserts are distributed during workshops and events. The new versions¹ are also directly available on the D4Science-II website.

D4Science-II Flyer

Project participants:

United Kingdom: BDM-ISTRATH
 Switzerland: CERN
 Hungary: 4D SOFT
 France: ERICIM
 Greece: NIKUA
 Italy: CNR-ISTI, ENG, FAD, TERRADUE
 The Philippines: The D4Science II
 Italy: ERICIM, GBE ERDM (ERICIM), Istituto di Scienza e Tecnologia dell'Informazione (ISTI), National and Kapodistrian University of Athens (NKUA), European Organization for Nuclear Research (CERN), Engineering - Informatica SpA (ENG), University of Strathclyde (BDM-ISTRATH), The Food and Agriculture Organization of the United Nations (FAO), 4D SOFT Számítástechnikai Kft (4D SOFT), Fishbase Information & Research Group Inc. (FIN), Terradue s.r.l. (TERRADUE)

For more information about the project, please visit our dedicated web site: www.d4science.eu or contact us at: info@d4science.eu

D4Science-II
Data Infrastructures ecosystem for science
www.d4science.eu

AquaMaps Insert

D4Science-II as a knowledge ecosystem

This is the Knowledge Ecosystem. Interoperable data infrastructures and repositories integrate seamlessly with services and scientific communities. The D4Science Infrastructure is reflective, acting as a virtual aggregator of e-infrastructure resources, while also offering these aggregate resources back to participating e-Infrastructures.

D4Science-II is a European e-infrastructure project that supports research by creating virtual communities in which scientists from different disciplines and locations share access to data, applications, processing and communications.

D4Science-II is co-funded by the European Commission's Seventh Framework Programme for Research and Technological Development. It builds upon the work of the DILIGENT and D4Science projects.

DILIGENT
A test bed for the management of data, metadata, software and hardware resources for the provision of integrated digital library services over grid-enabled infrastructures.
Software: gCube 0.9
Level: Testbed
October 2004 - November 2007

D4SCIENCE
A test bed for the management of data, metadata, software and hardware resources for the provision of integrated digital library services over grid-enabled infrastructures.
Software: gCube 1.0
Level: Testbed
January 2008 - December 2008

D4SCIENCE-II
An ecosystem of infrastructures that share resources across technological, administrative and disciplinary barriers using shared standards and interoperable interfaces.
Software: gCube 2
Level: Production
October 2009 - September 2011

Virtual Research Environments (VREs)
VREs offer on-demand grid-based services to users with no development or maintenance costs. By exploiting the larger information space available through the ecosystem the D4Science e-infrastructure will support ever more powerful VREs.

Through these VREs D4Science-II will address the needs of five scientific scenarios for which interacting with a knowledge ecology and sharing across e-infrastructure plays a fundamental role in enhancing the quality and efficiency of research.

INSPIRE
This next-generation High Energy Physics information system empowers scientists with innovative tools to promote successful research.
Advanced dynamic services for more than 30 000 high energy physicists using the powerful features of gCube.
Parallel full-text indexing of large document sets. Optical Character Recognition, offered as a generic service.
Computationally intensive processing of bibliographic information to provide better metrics on high-energy physics literature.
Extends INSPIRE capabilities and services.

DRIVER
Enriches DRIVER repository with resources hosted on D4Science II infrastructure and vice versa, allowing:
- Cross infrastructure information retrieval and access
- Objective-driven content transformation for DRIVER documents
- Other objectives such as bibliometric analysis, author disambiguation and massive parallel grid and cloud processing of DRIVER repository documents.
The results are indirectly accessible through the DRIVER Repository and gCube portals.

AQUAMAPS
Produces maps calculated by fitting geo-referenced species occurrence data with the known physicochemical distribution parameters of the aquatic milieu.
Combines data obtained from multiple sources with the computational capability of the Grid to generate more precise maps.
Exploited by the AquaMaps portal and by other infrastructures that take advantage of generated predictive data.

FCPPS (Fishery Country Profiles Production System)
Management of collaborative efforts for the development of comprehensive FAO Fishery Country Profiles.
Facilitates requested data sharing across fishery, biology, commodity, and taxonomic information systems.
Maintains provenance of intermediate data generated during the process of producing country reports.
Requires on-demand access to computational resources, retrieval and access to heterogeneous shared data and storage and management of compound dynamically generated information objects.
Serves FAO Fisheries and Aquaculture Department.

ICIS (Integrated Capture Information System)
Improves species occurrence data in synergy with AquaMaps.
Allows the inclusion of vessel monitoring system catch data at extremely high resolution to improve aquatic resource assessments.
Need for interoperability with different data sources, data harmonization and large on-demand processing capabilities.
A community of regional fishery bodies in charge of monitoring and managing fisheries and aquatic resources in the world's oceans.

gCUBE

Driver Insert

DRIVER VRE

The DRIVER Virtual Research Environment (VRE) enhances and exploits the interoperability capacities of gCube and DINE systems in order to allow the respective infrastructures to exploit resources that lie outside their administrative domain and functional capacity. DINE institutions seek content and computational resources that lie outside their administrative and functional domain.

DINE System Institutions:

- The Drive Repository: An infrastructure that harvests, aggregates and curates approximately 2,140,000 documents in 260 repositories from 36 countries.
- The OpenABE repository: A pilot infrastructure supported by the European Commission, to promote and strengthen OpenAccess initiatives.

Objectives:

- Functional enhancement of DINE institutions via additional services operating on gCube computational and storage resources.
- Mutual exposure of content hosted on gCube / Driver to each other's communities.

Indirect beneficiaries of DRIVER VRE are scientists, researchers, national community and European repository portals.

DRIVER VRE is a project of the DRIVER VRE consortium, which is a consortium of research institutions and research centres from 36 countries, including the following:

- The Drive Repository: An infrastructure that harvests, aggregates and curates approximately 2,140,000 documents in 260 repositories from 36 countries.
- The OpenABE repository: A pilot infrastructure supported by the European Commission, to promote and strengthen OpenAccess initiatives.

DRIVER VRE is a project of the DRIVER VRE consortium, which is a consortium of research institutions and research centres from 36 countries, including the following:

- The Drive Repository: An infrastructure that harvests, aggregates and curates approximately 2,140,000 documents in 260 repositories from 36 countries.
- The OpenABE repository: A pilot infrastructure supported by the European Commission, to promote and strengthen OpenAccess initiatives.



gCube Insert

gCUBE

gCube is a software platform to enable infrastructures for the creation, hosting and maintenance of dynamic virtual environments capable of satisfying complex research and collaboration needs of distributed scientific communities organized in Virtual Organisations.

gCube extends the infrastructure concept by enabling the organization of community resources into ordered, shared Virtual Research Environments, effectively promoting sharing and collaboration among scientific and administrative domains, while enforcing policies for resource exploitation. With a rich set of services and components it supports the full cycle of knowledge generation and sharing and enables the exchange of data and information within and across disciplinary and technological boundaries.

gCube architecture follows the Service Oriented Paradigm. It is logically organized in 4 layers, populated by 4 service Groups while an underlying layer of outside software provides fundamental facilities to service implementation and execution.

The lower layer is formed by gCore Framework (gCF) and gCube Hosting Nodes (gHN); gCF is an extensible software development framework that

hosts the implementation of software services that interact and integrate with the gCube infrastructures. The gHN is the environment that hosts gCube services on a system and essentially acts as the gateway of a computer resource (cpu, storage etc) into a gCube-powered site/infrastructure.



The service logical level is formed by the **Informational, Infrastructural and Organizational Services** that cover the areas of:

- Infrastructure Management, Marketing and Self-reorganization
- Virtual Research Environment Management
- Virtual Organization and Security Support Services
- Process Execution

This service group, along with the Business software, are collectively called the "enabling layer".

Two service groups related to data, information and knowledge management build on top of the enabling component, **Informational and Organizational Services**, cover aspects such as:

- Storage Provision
- Collection, Content, Metadata and Annotation Management
- Archive Ingest
- Content and Metadata Interconversion
- Ontology Management

Informational Services enable the efficient retrieval of information managed in Information Organization Services, by providing services such as:

- Metadata Ingesting
- Content Ingesting
- Personalization
- Distributed Information Retrieval (Content Source Description & Selection and Data Fusion)
- Search

On the logical top, **Operational Services** exist, consisting of the following types of components:

- Application Support Layer
- Portal
- User Profiles
- Administrative Portals
- Desktop clients

Beside these 4 service groups, **Operational Services** can be implemented on top or beside other gCube services by extending or integrating existing technology.



POSTERS

gCube Posters

D4Science-II Generic Poster

Scientific Infrastructures Ecosystem Interoperability based on the gCube PE2ng

The gCube Process Execution Engine (PE2ng) bridges the gaps among different e-infrastructures and their users by allowing scientists to exploit computational resources via higher level constructs and in a uniform manner.

Through a layered architecture PE2ng decouples business domain and infrastructure specific logic from the core "execution" functionality, allowing the latter to be applied to a multitude of use cases yet avoiding sub-optimal compromises imposed by strictly agnostic solutions.

The key features of the PE2ng framework include:

- Invocation of a wide range of logic components: SOAP and REST WebServices, Shell Scripts, Executable Binaries, POJOs, etc
- Control and monitor the execution of a processing flow
- Staging of data among different storage providers
- Streaming data among computation elements
- Expressive and powerful execution plan language
- Extensibility via providers for integration with different environments (storage, registry, security)

The key features of the PE2ng framework include:

- Invocation of a wide range of logic components: SOAP and REST WebServices, Shell Scripts, Executable Binaries, POJOs, etc
- Control and monitor the execution of a processing flow
- Staging of data among different storage providers
- Streaming data among computation elements
- Expressive and powerful execution plan language
- Extensibility via providers for integration with different environments (storage, registry, security)

Adaptors: The gateway to infrastructure interoperability in PE2ng

Through Adaptors, PE2ng allows integration of multi-infrastructure workflows.

PE2ng Adaptors operate on a specific third party language which they translate into native constructs, allowing for the creation of complex workflows that exploit several diverse technologies deployed on different infrastructures.

PE2ng extensible interoperability toolkit:

- Adaptors
- Environment providers
- Input/output filters
- Execution element flexibility

PE2ng and the INSPIRE case

INSPIRE is the next-generation High Energy Physics (HEP) information system, which empowers scientists with innovative tools to promote successful research. It is based on Invenio digital library technology [1] developed at CERN, and provides access to almost one million records. INSPIRE needs advanced computing services to provide enhanced functionality to the worldwide community of over 30,000 scientists.

Document OCR: Massive Optical Character Recognition (OCR) that enables the textual analysis within a large collection of scanned documents. OCR'ing jobs are executed using the JDL Adaptor and gLite Adaptor.

Full-text Indexing: Indexing of full-texts enables efficient searching in big collections. Combined with OCR, it allows full-text searching in the entire HEP literature, including scanned papers within a fraction of a second. Full-text indexing jobs are executed in parallel on Hadoop clusters, interfaced through the PE2ng Hadoop Adaptor.

PE2ng and the AquaMaps case

AquaMaps is an application tailored to predict global distributions of marine species initially designed for marine mammals and subsequently generalised to marine species [2], that generates color-coded species range maps using a half-degree latitude and longitude blocks.

gCube Ecological Modeling Suite: The AquaMaps execution is based on this modelling suite which allows the extrapolation of known species occurrences to determine environmental envelopes (species tolerances) and to predict future distributions by matching species tolerances against local environmental conditions (e.g. climate change and sea pollution). The execution of this complex workflow is out-sourced to PE2ng.

Very large volume of input and output data
Around 7,000 species;
HSPEC native range 56,468,301
HSPEC suitable range 114,989,360

Very large number of computations
One multispecies map computed on 6,188 half degree cells (over 170k) and 2,540 species requires 125 millions computations [3]

(1) CERN Document Server Invenio (http://indiana.cern.ch/invenio)
(2) G. E. Agbayan, FishBase Project/ICR/FISH WP1, WWF/Fish Center
(3) E. E. Agbayan, FishBase Project/ICR/FISH WP1, WWF/Fish Center

DATA INFRASTRUCTURES ECOSYSTEM FOR SCIENCE

"Knowledge Ecology is knowledge generated by analysing information from diverse sources for a more effective evaluation"
Thomas Garnett, Second GRL 2020 Workshop, Pisa (Italy), March 2008

The D4Science enabling technology is gCube (www.gcube-system.org)

The D4Science II aims to provide mechanisms for interoperability with other data e-infrastructures, thus creating the core of an e-Infrastructure Ecosystem. Initially, this core will include:

- AQUAMAPS
- DRIVER
- FAO FISHERIES
- GENESI-DR
- INSPIRE

Virtual Research Environments (VREs)

Virtual Research Environments (VREs) offer services to their users without incurring high development and maintenance costs.

By exploiting the larger information space available through the ecosystem the D4Science e-infrastructure will be able to support more powerful VREs.

Through these VREs D4Science-II will address the needs of five scientific scenarios for which interacting with a Knowledge Ecology and sharing across e-infrastructures plays a fundamental role in enhancing the quality and efficiency of research.

INSPIRE VRE

- Mining of bibliometric data and calculation of hybrid metrics on the entire corpus of the High Energy Physics literature
- Needs dynamic service deployment and on-demand processing
- Exploited by the INSPIRE service

DRIVER VRE

- Metadata brokering, data transformations and ontology-based mapping
- Support for the creation, storage and management of publications with the possibility to link content available under appropriate policies in the ecosystem
- Mining of bibliometric data and calculation of hybrid metrics on the DRIVER accessible corpus
- Requires on-demand processing, access to primary data and to shared tools
- Made available through the DRIVER e-infrastructure portal

AQUAMAPS VRE


- Produces maps that result from fitting geo-referenced species occurrence data with known distribution of physico-chemical parameters of the aquatic milieu
- Exploits data obtained from multiple sources and the computational capability of the Grid to generate more precise maps
- Exploited by the AquaMaps portal and by other infrastructures that can take advantage of the generated predictive data

FAO FISHERIES VRE


- Management of collaborative efforts for the development of comprehensive FAO Fishery Country Profiles report
- Facilitates the requested data sharing across fishery, biology, commodity, and taxonomic information systems
- Maintains track of intermediate data generated when producing country reports, since now discarded
- Requires on-demand access to computational resources, retrieval and access to heterogeneous shared data, storage and management of dynamically generated compound information objects

ICIS VRE

- Management of a repository of harmonized statistical data sources for improving catch statistics, in synergy with Aquamaps
- Allows the inclusion of Vessel Monitoring System Catch data at extremely high resolution to improve aquatic resource assessments.
- Interoperable with different data sources, data harmonization, large on-demand processing capabilities.
- Uses data from Regional Fisheries Bodies and provides harmonization and mapping features for capture statistics.




D4SCIENCE II
www.d4science.eu
info@d4science.eu



ICES VIRTUAL RESEARCH ENVIRONMENT
Integrated Catch Information System

ICES is a FAO initiative in response to the United Nations General Assembly (UNGA) request to generate fishery statistics that distinguish catches within and outside the EEZs, and to the Coordinating Working Party on fishery statistics (CWP) to improve the quality of global catch statistics.

Implemented as Virtual Research Environment (VRE) on the D4Science infrastructure, ICIS exploits VRE facilities to manage complex collaborative efforts and assemble an advanced mix of Information Technology (IT). The ICIS VRE will allow users to load and store their catch statistical sources on the D4Science infrastructure, to harmonise them with other data sources, to analyze data across domains, and to present results in reports, tables, graphs or maps. The main objective of ICIS is to improve global catch statistics quality.



ICES users: Fishery biologists, data analysts, and policymakers.

D4Science is developing the ICIS VRE to achieve the following Objectives

- Data Harmonization:** make data from different sources comparable, e.g. to merge various sources in unified statistical time-series analysis
- Spatial Capture Re-allocation:** improve spatial precision in capture statistics by applying spatial logic
- Interoperability:** establish seamless links to external repositories
- SDMX Support:** support data generation in the statistical data exchange format SDMX, and store and expose these data in a secure but open infrastructure.

Data-harmonization of fisheries statistics

- ICES offers a variety of **transformation services** to harmonize data from different sources
- Data can be imported as csv files, and transformations to e.g. **SDMX** formats are supported
- ICES offers reference data management facilities driven by **semantic web technologies**
- A semantic web knowledge base** manages relations between concepts

Spatial re-allocation on the grid

- D4Science has pioneered a **grid implementation of geospatial open standards**, such as GeoServer, offering tailored and high-performance solutions: In mapping and geo-spatial analysis
- ICES data can be displayed and saved in a variety of **mapping formats**
- Capture statistics can be combined with data from other domains, such as species distribution maps from AquaMaps
- Data can be generated in a variety of **exportable formats**

Interoperability

- D4Science offers a range of **collaborative services** to e.g. share huge datasets, edit reports, import data from other VREs, and publish to a variety of formats, all in one e-infrastructure
- Data can be imported from Webservices available in other e-infrastructures, feeds, files, or programmatic interfaces
- Concurrent editing, user-to-user forwarding of datasets, workflow support and metadata management all enhance the interoperable features of the VREs

SDMX Development

- SDMX is an XML statistical data exchange standard supported by e.g. the EU, the WorldBank, OECD, and FAO. It also covers the management of metadata and the discovery of data-sets.
- FAO-FI supports development of SDMX capabilities in D4Science to provide SDMX generating and storage capabilities to institutions that lack resources to implement the (extensive) standard
- Open SDMX is a recent FAO initiative where JAVA developers can contribute to the implementation of additional SDMX features on the grid

Introduction to D4Science

- D4Science is an FP7 EU funded Information Technology initiative aimed at providing collaborative scientific environments
- D4Science is currently supporting the development of Virtual Research Environments for reporting, statistical data analysis and mapping, earth observation science digital repositories, and virtual libraries
- Virtual research environments are collections of data and services, bundled in a secure environment on a grid infrastructure. The grid infrastructure reduces hardware and software maintenance costs, and offers a wide range of information and user management capabilities

Links and Further Info

The Integrated Capture Information System (ICIS) is developed with support from the Japanese Government (FAO project GCP/INT/099/JPN) and the EU FP7-D4Science project.

The Japanese funds are used to develop data harmonization and spatial re-allocation logic, while the D4Science funds are used to develop interoperability and the underlying IT infrastructure

More information can be found on the D4Science website: <http://www.d4science.org/>





VIDEOS: an effective means to communicate relevant and complex information to D4Science-II website visitors. During the full project period, there were three videos were filmed:

- Fishery Country Profiles Production System (FCPPS) presented by Marc Taconet highlighted a specific use case for the fisheries community that has been implemented within the project’s grid infrastructure;
- Integrated Capture Information System (ICIS) presented by Anton Ellenbroek presented a second use case for the fisheries community that has been implemented within the project’s grid infrastructure.
- D4Science-II video about GRDI 2020 meets the D4Science Project;

Fishery Country Profiles Production System (FCPPS)



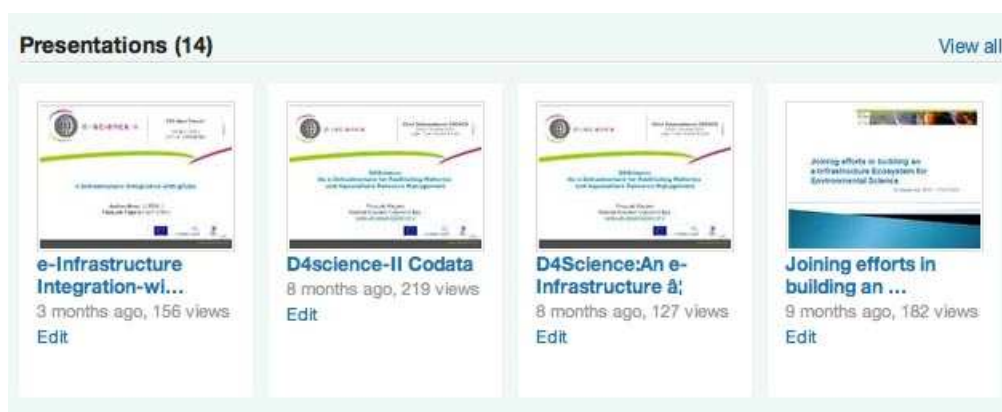
Integrated Capture Information System (ICIS) video



- Blip TV⁷ and YouTube (as a secondary channel) are used to add D4Science-II project videos. These videos create a personal, interactive experience between the project and the audience, engaging them and keeping visitors on the site for a longer time.

⁷ <http://blip.tv/file/1912552>

SLIDES: Public PowerPoint files prepared during the reporting period were added to the popular presentation sharing site Slideshare and are linked from the D4Science-II web site



PHOTOS: D4Science–II photos, posted on the D4Science Flickr Gallery⁴, illustrate project participation at events and meetings. All the pictures are organized using tags enabling search related to particular topics such as event name, subject matter, and place.



D4S-II WEBSITE:

The D4Science-II public web site is an important and versatile tool to communicate project progress and events to existing and potential community members. As the main dissemination vehicle it includes all publications, news, journal articles, presentations at conferences as well as core documents such as project deliverables. Project members are collaborating on the update of the web site by proposing new stories about the project in order to raise awareness among the widest possible audience.

The D4Science-II blog is the most frequently updated part of the website and it is among the most frequented pages of the site (see Table 4). All new information is shared here with project participants and user communities.

New areas were also added. There is now a general page with an explanation regarding the projects collaborating with D4Science-II and links to each of these projects. A new block on the right side of the page has been created in order to easily reach the fifteen sub pages that detail each collaboration. Links are illustrated with logos to increase brand recognition.

Collaborating Projects

The screenshot shows the D4Science-II website. The main content area is titled "Collaborating projects" and features a diagram with arrows connecting various project acronyms. The diagram includes: EMI, gMan, VENUS-C, #D4Lite&ALife, DRIVER-II, GENESI-DEC, INSPIRE, OBIS, OpenAIRE, SDMX, BELIEF-II, and [all projects].

Collaborating projects

- EMI
- gMan
- VENUS-C
- #D4Lite&ALife
- DRIVER-II
- GENESI-DEC
- INSPIRE
- OBIS
- OpenAIRE
- SDMX
- BELIEF-II
- ...
- [all projects]

Under the Documentation menu a new “Presentations”⁸ section has been created providing direct access via an embedded gadget that takes visitors directly to the D4Science-II Slideshare⁹ account.

D4Science-II Slideshare presentations

⁸ <http://www.d4science.eu/presentations>

⁹ <http://www.slideshare.net/d4science>

Slideshare Presentations



View all presentations

A monthly news digest¹⁰ has also been introduced. It is a summary of the news published on the D4Science-II blog. It is sent to all project contacts collected during different events – the largest set those who attended the D4Science World User Meeting¹¹ in 2009.

Monthly News Digest



If you can not visualize this e-mail, please click [here](#)

The D4Science-II monthly news digest delivers news and other information published on the web site during the course of the month.

June 2011

INSPIRE Community Applications
A set of prototype gCube components related to INSPIRE use cases was created by the NKUA team in cooperation with INSPIRE.
[read more...](#)

Presentation of gCube system evolution and its architecture
On Saturday the 11th of June 2011, the MADgIK Lab, the NKUA group that participates D4Science-II project, has organized an event inviting a representative group of Architects and Managers of the SQL Server RDBMS, from Microsoft Corp.
[read more...](#)

Presentation of the Development of the VTC Virtual Research Environment (VTC) to NAFO STACREC Science Council Meeting
The use of Vessel Transmitted Information by fishery scientists is rapidly growing. The continuous improvement of their detail and volume has opened exciting new possibilities for scientist working in the fields of marine biology, ecology and environmental sciences.
[read more...](#)

If you do not wish to receive our monthly news digest please let us know by sending us an [e-mail](#)

¹⁰ <http://www.d4science.eu/node/675>

¹¹ <http://www.d4science.eu/worldusermeeting/>

PAPERS: 14 D4Science-II papers were published. Their electronic versions² are also available on a dedicated section of the D4Science-II web site. (See Table A1: List of scientific publications) or the webpage [dedicated listing all papers](#)

EGEE JOINT PUBLICATIONS

- [Building Scientific Workflows for the Fisheries and Aquaculture Management Community based on Virtual Research Environments](#) 5th EGEE User Forum, April 2010

Appendix C - Report on societal implications

Replies to the following questions will assist the European Commission to obtain statistics and indicators on societal and socio-economic issues addressed by projects. The questions are arranged in a number of key themes. As well as producing certain statistics, the replies will also help identify those projects that have shown a real engagement with wider societal issues, and thereby identify interesting approaches to these issues and best practices. The replies for individual projects will not be made public.

A General Information <i>(completed automatically when Grant Agreement number is entered.</i>		
Grant Agreement Number:	239019	
Title of Project:	D4Science-II	
Name and Title of Coordinator:	Jessica Michel Project Coordinator (ERCIM)	
B Ethics		
1. Did your project undergo an Ethics Review (and/or Screening)?	Yes	No
If yes: have you described the progress of compliance with the relevant Ethics Review/Screening Requirements in the frame of the periodic/final project report		
2. Please indicate whether your project involved any of the following issues (tick box) :	YES	
RESEARCH ON HUMANS		
• Did the project involve children?		
• Did the project involve patients ?		
• Did the project involve persons not able to give consent?		
• Did the project involve adult healthy volunteers?		
• Did the project involve Human Genetic Material?		
• Did the project involve Human biological samples?		
• Did the project involve Human data collection?		
RESEARCH ON HUMAN EMBRYO/FOETUS		
• Did the project involve Human Embryos?		
• Did the project involve Human Foetal Tissue / Cells?		
• Did the project involve Human Embryonic Stem Cells?		
• Did the project on human Embryonic Stem Cells involve cells in culture?		
• Did the project on human Embryonic Stem Cells involve the derivation of cells from Embryos?		
PRIVACY		
• Did the project involve processing of genetic information or personal data (eg. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?		
• Did the project involve tracking the location or observation of people?		
RESEARCH ON ANIMALS		
• Did the project involve research on animals?		
• Were those animals transgenic small laboratory animals?		
• Were those animals transgenic farm animals?		
• Were those animals cloned farm animals?		
• Were those animals non-human primates?		
RESEARCH INVOLVING DEVELOPING COUNTRIES		
• Did the project involved the use of local resources (genetic, animal, plant etc)		
• Was the project of benefit to local community (capacity building ie access to healthcare, education etc)		

DUAL USE		
<ul style="list-style-type: none"> • Research having direct military use 		
<ul style="list-style-type: none"> • Research having the potential for terrorist abuse 		
C Workforce Statistics		
3 Workforce statistics for the project: Please indicate in the table below the number of people who worked on the project (on a headcount basis).		
Type of Position	Number of Women	Number of Men
Scientific Coordinator	1	
Work package leader		
Experienced researcher (i.e. PhD holders)	4	26
PhD Students	2	6
Other	13	23
4 How many additional researchers (in companies and universities) were recruited specifically for this project?		7
Of which, indicate the number of men:		7

D Gender Aspects		
5 Did you carry out specific Gender Equality Actions under the project?	<input checked="" type="radio"/> Yes <input type="radio"/> No	<input type="radio"/> Yes <input checked="" type="radio"/> No
6 Which of the following actions did you carry out and how effective were they?		
<input type="checkbox"/> Design and implement an equal opportunity policy <input type="checkbox"/> Set targets to achieve a gender balance in the workforce <input type="checkbox"/> Organise conferences and workshops on gender <input type="checkbox"/> Actions to improve work-life balance <input checked="" type="checkbox"/> Other:	Not at all effective <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Very effective <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Publication of the D4Science-II Gender statement on the D4Science website http://www.d4science.eu/statement		
7 Was there a gender dimension associated with the research content – i.e. wherever people were the focus of the research as, for example, consumers, users, patients or in trials, was the issue of gender considered and addressed?		
<input type="radio"/> Yes- please specify 		
<input checked="" type="radio"/> No		
E Synergies with Science Education		
8 Did your project involve working with students and/or school pupils (e.g. open days, participation in science festivals and events, prizes/competitions or joint projects)?		
<input type="radio"/> Yes- please specify		
<input checked="" type="radio"/> No		
9 Did the project generate any science education material (e.g. kits, websites, explanatory booklets, DVDs)?		
<input checked="" type="radio"/> Yes- please specify (Training material, explanatory booklets)		
<input type="radio"/> No		
F Interdisciplinarity		
10 Which disciplines (see list below) are involved in your project?		
<input checked="" type="checkbox"/> Main discipline ¹² :		
<input checked="" type="checkbox"/> Associated discipline ¹² :		
		<input checked="" type="checkbox"/> Associated discipline ¹² :

¹² Insert number from list below (Frascati Manual)

G Engaging with Civil society and policy makers							
11a	Did your project engage with societal actors beyond the research community? (if 'No', go to Question 14)	<input checked="" type="checkbox"/> <input type="checkbox"/>	Yes No				
11b	If yes, did you engage with citizens (citizens' panels / juries) or organised civil society (NGOs, patients' groups etc.)? <input type="checkbox"/> No <input checked="" type="checkbox"/> Yes- in determining what research should be performed <input checked="" type="checkbox"/> Yes - in implementing the research <input checked="" type="checkbox"/> Yes, in communicating /disseminating / using the results of the project						
11c	In doing so, did your project involve actors whose role is mainly to organise the dialogue with citizens and organised civil society (e.g. professional mediator; communication company, science museums)?	<input type="checkbox"/> <input checked="" type="checkbox"/>	Yes No				
12	Did you engage with government / public bodies or policy makers (including international organisations) <input type="checkbox"/> No <input checked="" type="checkbox"/> Yes- in framing the research agenda <input checked="" type="checkbox"/> Yes - in implementing the research agenda <input checked="" type="checkbox"/> Yes, in communicating /disseminating / using the results of the project						
13a	Will the project generate outputs (expertise or scientific advice) which could be used by policy makers? <input checked="" type="checkbox"/> Yes – as a primary objective (please indicate areas below- multiple answers possible) <input type="checkbox"/> Yes – as a secondary objective (please indicate areas below - multiple answer possible) <input type="checkbox"/> No						
13b	If Yes, in which fields? <table border="1" style="width:100%; border-collapse: collapse;"> <tr> <td style="width:25%; vertical-align: top;"> Agriculture Audiovisual and Media Budget Competition Consumers Culture Customs Development Economic and Monetary Affairs Education, Training, Youth Employment and Social Affairs </td> <td style="width:25%; vertical-align: top;"> Energy Enlargement Enterprise Environment External Relations External Trade Fisheries and Maritime Affairs Food Safety Foreign and Security Policy Fraud Humanitarian aid </td> <td style="width:25%; vertical-align: top;"> Human rights Information Society Institutional affairs Internal Market Justice, freedom and security Public Health Regional Policy Research and Innovation Space Taxation Transport </td> <td style="width:25%;"></td> </tr> </table>			Agriculture Audiovisual and Media Budget Competition Consumers Culture Customs Development Economic and Monetary Affairs Education, Training, Youth Employment and Social Affairs	Energy Enlargement Enterprise Environment External Relations External Trade Fisheries and Maritime Affairs Food Safety Foreign and Security Policy Fraud Humanitarian aid	Human rights Information Society Institutional affairs Internal Market Justice, freedom and security Public Health Regional Policy Research and Innovation Space Taxation Transport	
Agriculture Audiovisual and Media Budget Competition Consumers Culture Customs Development Economic and Monetary Affairs Education, Training, Youth Employment and Social Affairs	Energy Enlargement Enterprise Environment External Relations External Trade Fisheries and Maritime Affairs Food Safety Foreign and Security Policy Fraud Humanitarian aid	Human rights Information Society Institutional affairs Internal Market Justice, freedom and security Public Health Regional Policy Research and Innovation Space Taxation Transport					
13c	If Yes, at which level? <input checked="" type="checkbox"/> Local / regional levels <input checked="" type="checkbox"/> National level <input checked="" type="checkbox"/> European level <input checked="" type="checkbox"/> International level						

H Use and dissemination			
14	How many Articles were published/accepted for publication in peer-reviewed journals?	0	
To how many of these is open access¹³ provided?		-	
How many of these are published in open access journals?		-	
How many of these are published in open repositories?		-	
To how many of these is open access not provided?		-	
Please check all applicable reasons for not providing open access:		-	
<input type="checkbox"/> publisher's licensing agreement would not permit publishing in a repository <input type="checkbox"/> no suitable repository available <input type="checkbox"/> no suitable open access journal available <input type="checkbox"/> no funds available to publish in an open access journal <input type="checkbox"/> lack of time and resources <input type="checkbox"/> lack of information on open access <input type="checkbox"/> other:			
15	How many new patent applications ('priority filings') have been made? <i>("Technologically unique": multiple applications for the same invention in different jurisdictions should be counted as just one application of grant).</i>	0	
16	Indicate how many of the following Intellectual Property Rights were applied for (give number in each box).	Trademark	0
		Registered design	0
		Other	0
17	How many spin-off companies were created / are planned as a direct result of the project?	0	
<i>Indicate the approximate number of additional jobs in these companies:</i>			
18	Please indicate whether your project has a potential impact on employment, in comparison with the situation before your project:		
	<input type="checkbox"/> Increase in employment, or <input type="checkbox"/> Safeguard employment, or <input type="checkbox"/> Decrease in employment, <input checked="" type="checkbox"/> Difficult to estimate / not possible to quantify	<input type="checkbox"/> In small & medium-sized enterprises <input type="checkbox"/> In large companies <input type="checkbox"/> None of the above / not relevant to the project <input type="checkbox"/>	
19	For your project partnership please estimate the employment effect resulting directly from your participation in Full Time Equivalent (FTE = one person working fulltime for a year) jobs:	<i>Indicate figure:</i>	
	1 FTE was made possible through the project	<input type="checkbox"/>	
	<i>Difficult to estimate / not possible to quantify</i>		

¹³ Open Access is defined as free of charge access for anyone via the internet.

I Media and Communication to the general public			
20	<p>As part of the project, were any of the beneficiaries professionals in communication or media relations?</p> <p><input type="radio"/> Yes <input checked="" type="radio"/> No</p>		
21	<p>As part of the project, have any beneficiaries received professional media / communication training / advice to improve communication with the general public?</p> <p><input type="radio"/> Yes <input checked="" type="radio"/> No</p>		
22	<p>Which of the following have been used to communicate information about your project to the general public, or have resulted from your project?</p> <table border="0" style="width: 100%;"> <tr> <td style="width: 50%; vertical-align: top;"> <input checked="" type="checkbox"/> Press Release <input type="checkbox"/> Media briefing <input type="checkbox"/> TV coverage / report <input type="checkbox"/> Radio coverage / report <input checked="" type="checkbox"/> Brochures /posters / flyers <input checked="" type="checkbox"/> DVD /Film /Multimedia </td> <td style="width: 50%; vertical-align: top;"> <input type="checkbox"/> Coverage in specialist press <input type="checkbox"/> Coverage in general (non-specialist) press <input type="checkbox"/> Coverage in national press <input type="checkbox"/> Coverage in international press <input checked="" type="checkbox"/> Website for the general public / internet <input checked="" type="checkbox"/> Event targeting general public (festival, conference, exhibition, science café) </td> </tr> </table>	<input checked="" type="checkbox"/> Press Release <input type="checkbox"/> Media briefing <input type="checkbox"/> TV coverage / report <input type="checkbox"/> Radio coverage / report <input checked="" type="checkbox"/> Brochures /posters / flyers <input checked="" type="checkbox"/> DVD /Film /Multimedia	<input type="checkbox"/> Coverage in specialist press <input type="checkbox"/> Coverage in general (non-specialist) press <input type="checkbox"/> Coverage in national press <input type="checkbox"/> Coverage in international press <input checked="" type="checkbox"/> Website for the general public / internet <input checked="" type="checkbox"/> Event targeting general public (festival, conference, exhibition, science café)
<input checked="" type="checkbox"/> Press Release <input type="checkbox"/> Media briefing <input type="checkbox"/> TV coverage / report <input type="checkbox"/> Radio coverage / report <input checked="" type="checkbox"/> Brochures /posters / flyers <input checked="" type="checkbox"/> DVD /Film /Multimedia	<input type="checkbox"/> Coverage in specialist press <input type="checkbox"/> Coverage in general (non-specialist) press <input type="checkbox"/> Coverage in national press <input type="checkbox"/> Coverage in international press <input checked="" type="checkbox"/> Website for the general public / internet <input checked="" type="checkbox"/> Event targeting general public (festival, conference, exhibition, science café)		
23	<p>In which languages are the information products for the general public produced?</p> <table border="0" style="width: 100%;"> <tr> <td style="width: 50%; vertical-align: top;"> <input type="checkbox"/> Language of the coordinator <input type="checkbox"/> Other language(s) </td> <td style="width: 50%; vertical-align: top;"> <input checked="" type="checkbox"/> English </td> </tr> </table>	<input type="checkbox"/> Language of the coordinator <input type="checkbox"/> Other language(s)	<input checked="" type="checkbox"/> English
<input type="checkbox"/> Language of the coordinator <input type="checkbox"/> Other language(s)	<input checked="" type="checkbox"/> English		

Question F-10: Classification of Scientific Disciplines according to the Frascati Manual 2002 (Proposed Standard Practice for Surveys on Research and Experimental Development, OECD 2002):

FIELDS OF SCIENCE AND TECHNOLOGY

1. NATURAL SCIENCES

- 1.1 Mathematics and computer sciences [mathematics and other allied fields: computer sciences and other allied subjects (software development only; hardware development should be classified in the engineering fields)]
- 1.2 Physical sciences (astronomy and space sciences, physics and other allied subjects)
- 1.3 Chemical sciences (chemistry, other allied subjects)
- 1.4 Earth and related environmental sciences (geology, geophysics, mineralogy, physical geography and other geosciences, meteorology and other atmospheric sciences including climatic research, oceanography, vulcanology, palaeoecology, other allied sciences)
- 1.5 Biological sciences (biology, botany, bacteriology, microbiology, zoology, entomology, genetics, biochemistry, biophysics, other allied sciences, excluding clinical and veterinary sciences)

2 ENGINEERING AND TECHNOLOGY

- 2.1. Civil engineering (architecture engineering, building science and engineering, construction engineering, municipal and structural engineering and other allied subjects)
- 2.2. Electrical engineering, electronics [electrical engineering, electronics, communication engineering and systems, computer engineering (hardware only) and other allied subjects]

2.3. Other engineering sciences (such as chemical, aeronautical and space, mechanical, metallurgical and materials engineering, and their specialised subdivisions; forest products; applied sciences such as geodesy, industrial chemistry, etc.; the science and technology of food production; specialized technologies of interdisciplinary fields, e.g. systems analysis, metallurgy, mining, textile technology and other applied subjects)

3. MEDICAL SCIENCES

3.1 Basic medicine (anatomy, cytology, physiology, genetics, pharmacy, pharmacology, toxicology, immunology and immunohaematology, clinical chemistry, clinical microbiology, pathology)

3.2 Clinical medicine (anaesthesiology, paediatrics, obstetrics and gynaecology, internal medicine, surgery,

dentistry, neurology, psychiatry, radiology, therapeutics, otorhinolaryngology, ophthalmology)

3.3 Health sciences (public health services, social medicine, hygiene, nursing, epidemiology)

4. AGRICULTURAL SCIENCES

4.1 Agriculture, forestry, fisheries and allied sciences (agronomy, animal husbandry, fisheries, forestry, horticulture, other allied subjects)

4.2 Veterinary medicine

5. SOCIAL SCIENCES

5.1 Psychology

5.2 Economics

5.3 Educational sciences (education and training and other allied subjects)

5.4 Other social sciences [anthropology (social and cultural) and ethnology, demography, geography (human, economic and social), town and country planning, management, law, linguistics, political sciences, sociology, organisation and methods, miscellaneous social sciences and interdisciplinary, methodological and historical S1T activities relating to subjects in this group. Physical anthropology, physical geography and psychophysiology should normally be classified with the natural sciences].

6. HUMANITIES

6.1 History (history, prehistory and history, together with auxiliary historical disciplines such as archaeology, numismatics, palaeography, genealogy, etc.)

6.2 Languages and literature (ancient and modern)

6.3 Other humanities [philosophy (including the history of science and technology) arts, history of art, art criticism, painting, sculpture, musicology, dramatic art excluding artistic "research" of any kind, religion, theology, other fields and subjects pertaining to the humanities, methodological, historical and other S1T activities relating to the subjects in this group]