

D4Science - Deploying Virtual Research Environments

by Donatella Castelli and Jessica Michel

The D4Science (Distributed colLaboratories Infrastructure on Grid-Enabled Technology for Science) project aims to continue the path that the GÉANT (a multi-gigabit pan-European data communications network), EGEE (Enabling Grids for E-science in Europe), and DILIGENT (A Digital Library Infrastructure on Grid Enabled Technology) projects have initiated towards establishing networking, Grid-based, and data-centric e-Infrastructures. These e-Infrastructures are expected to accelerate multidisciplinary research by overcoming several crucial barriers that stand in the way, primarily those related to heterogeneity, sustainability and scalability. In order to achieve this objective a D4Science production quality e-Infrastructure has been created, and will be progressively enriched and consolidated. It will provide facilities for creating Virtual Research Environments based on shared computational, data and service resources offered by many different providers like EGEE and large international organizations.

The D4Science project builds on the experience of a predecessor testbed project, DILIGENT (A Digital Library Infrastructure on Grid ENabled Technology). This was the first project to propose the integration of digital library and grid technology in order to create an innovative type of e-Infrastructure capable of supporting scientific cooperation based on the managed sharing of a variety of resources. The core notions of the solution proposed by DILIGENT are virtual organizations (modelling sets of users and resources and clearly defining what is shared, who is allowed to share, and the conditions under which sharing occurs), and virtual research environments (representing frameworks of applications, service and data sources dynamically identified to support cooperation processes). By implementing these two notions, the DILIGENT e-Infrastructure provides a framework for improving synergies across currently fragmented scientific communities, thus aiding research on a global scale.

User Communities

D4Science aims to consolidate and exploit the DILIGENT e-Infrastructure to build VREs that will serve the needs of broad scientific disciplines. In its first stage, the focus is on serving the domains of environmental monitoring, and fisheries and aquaculture resource management. The project's user communities – the European Space Agency (ESA), departments of the Food and Agriculture Organization of the United Nations (FAO), and the World Fish Center, a member of the Consultative Group on International Agricultural Research – will benefit greatly from access to shared resources. The e-Infrastructure will, for example, provide them with a powerful instrument for ecosystem modelling. This application requires satellite data (eg ocean colour and reef maps), climate data, hydrographic data and a large variety of other types of data (eg environmental reports, species taxonomies, time series etc), as well as specific, often computationally intensive tools for processing this data.



A large aquamarine-coloured phytoplankton bloom is shown stretching across the length of Ireland in the North Atlantic Ocean in this image, captured on 6 June 2006 by Envisat's Medium Resolution Imaging Spectrometer (MERIS), a dedicated ocean colour sensor able to identify plankton concentrations. Harmful algal blooms (HABs) is particularly dangerous to fish farms because the fish cannot flee affected areas. D4Science can, for example, provide the infrastructure for an early warning system of HABs to prevent fish farmers from losing an entire stock in a single day. Photo: European Space Agency.

Cooperation between these partners will also encourage the use of data streams from diverse scientific communities to perform socio-ecosystem modelling. The work that the two communities wish to undertake can be applied to the whole domain of biodiversity management (natural resources management), biodiversity conservation, and biodiversity exploitation as catch (mainly fisheries) and farming (both land and aquatic, ie aquaculture). The project will therefore bring together not only the typically separated communities of fisheries/aquaculture and environmental monitoring, but will also create a framework and a process to promote further cooperation of this type across other related communities. Technological Challenges D4Science addresses many important technological challenges. In particular, it will:

- Consolidate and enhance the technology which underpins the D4Science e-Infrastructure operation, namely the gCube framework. gCube, successfully deployed within the testbed developed by the DILIGENT project, reflects within its name a three-sided interpretation of the grid vision of resource sharing: sharing of computational resources, structured data and application services. As such, gCube embodies the defining characteristics of computational grids, data grids and virtual data grids. More precisely, it builds on gLite middleware (developed by the EGEE project) for managing distributed computations and unstructured data, includes dedicated services for managing data and metadata, and offers a novel approach for managing application services. Rather than interfacing with the infrastructure, the software that implements the services is literally handed over to it, so as to be transparently deployed across its constituent nodes according to functional constraints and quality-of-service requirements. This is genuinely ambitious and entirely novel: like computational resources and data before, application logic in gCube becomes a pervasive commodity within an infrastructure that abstracts over the physical location of its resources at any point in time. D4Science now consolidates and enhances the gCube services to reflect the shift in functional and QoS requirements, which marks the passage from a testbed to a production-level infrastructure.
- Make available and maintain a stable, reliable and fully operational e-Infrastructure that offers secure sharing and dynamic allocation of resources, where the notion of resources is extended to content and services in addition to computing and storage. By progressively integrating new sites and collecting feedback and requirements, the procedures and policies for managing and operating this infrastructure will also be clarified and consolidated. D4Science will thus become one of the first fully operational infrastructures for virtual data grids.
- Empower the mechanisms for the definition of virtual research environments (by introducing simplified procedures for resource discovery and selection) and their management (by implementing transparent bootstrapping, dynamic configuration and recovery mechanisms for stateful resources; more sophisticated load-balancing mechanisms during resource deployment and discovery; more flexible notification brokering mechanisms and update policies for resource change; multi-view monitoring mechanisms for resource administration; and finer-grained authentication and authorization policies for resource sharing).

D4Science is one of the main European Research e-Infrastructure projects, supported by the European Commission's Seventh Framework Programme for Research and Technological Development. ERCIM manages the project's administrative and financial activities, while ISTI-CNR leads the scientific coordination. The project started in January 2008 and will have a duration of two years.

It is expected that the D4Science e infrastructure will have a multiplicative benefit to many scientific fields, and will also act as a catalyst for the kind of cooperation and cross-fertilization among multiple communities that is necessary to address many of the grand challenges of science and society.

Links:

D4Science: <http://www.d4science.eu/>

gCube: <http://www.gcube-system.org/>

gLite Lightweight Middleware for Grid Computing: <http://glite.web.cern.ch/glite/>

EGEE: Enabling Grids for E-science in Europe: <http://public.eu-egee.org>

Please contact:

Donatella Castelli, project director
ISTI-CNR, Italy
E-mail: donatella.castelli@isti.cnr.it

Jessica Michel, administrative and financial director
ERCIM office

E-mail: jessica.michel@ercim.org