d4science

=====

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

Deliverable No DNA2.2a

Report on inter-projects coordination and collaboration

August 2008

SEVENTH FRAMEWORK PROGRAMME
Research Infrastructures

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

DOCUMENT INFORMATION

Project	
Project acronym:	D4Science
Project full title:	**DI**stributed col**L**aboratories **I**nfrastructure on **G**rid **EN**abled **T**echnology **4 Science**
Project start:	1st January 2008
Project duration:	24 months
Call:	INFRA-2007-1.2.2: Deployment of e-Infrastructures for scientific communities
Grant agreement no.:	212488
Document	
Deliverable number:	DNA2.2a
Deliverable title:	Report on inter-projects coordination and collaboration
Contractual Date of Delivery:	August 2008
Actual Date of Delivery:	18 November 2008
Editor(s):	Donatella Castelli
Author(s):	Donatella Castelli, Maria Bruna Baldacci
Contributor(s):	Roberta Faggian, Veronica Guidetti, Yves Jaques, George Kakaletris, and Fabio Simeoni
Reviewer(s):	Jessica Michel
Participant(s):	CNR, NKUA, ESA, FAO, CERN
Work package no.:	NA2
Work package title:	Scientific and Technological coordination
Work package leader:	CNR
Work package participants:	CNR, NKUA, CERN, ESA, FAO
Est. Person-months:	2.5
Distribution:	Public
Nature:	Report
Version/Revision:	1.0
Draft/Final	Final
Total number of pages: (including cover)	34
Keywords:	Inter-project collaboration, cross-fertilization, technological exploitation, networking

DISCLAIMER

This document contains description of the D4Science project findings, work and products. Certain parts of it might be under partner Intellectual Property Right (IPR) rules so, prior to using its content please contact the consortium head for approval. E-mail: info@d4science.research-infrastructures.eu

In case you believe that this document harms in any way IPR held by you as a person or as a representative of an entity, please do notify us immediately.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of D4Science consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 27 Member States of the Union. It is based on the European Communities and the member states cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors. (http://europa.eu.int/)

D4Science is a project partially funded by the European Union

LIST OF ABBREVIATIONS

AgMES	Agricultural Metadata Element Set
BELIEF	Bringing Europe's eLectronic Infrastructures to Expanding Frontiers
BHL	Biodiversity Heritage Library
CASPAR	Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval
CEN	European Committee for Standardization
CLARIN	Common Language Resources and Technology Infrastructure Network
CWP	Coordinated Working Party
D4Science	DIstributed colLaboratories Infrastructure on Grid Enabled Technology 4 Science
DCMI	Dublin Core Metadata Initiative
DILIGENT	DIgital Library Infrastructure on Grid Enabled Technology
DPE	Digital Preservation Europe
DRIVER	Digital Repository Infrastructure Vision for European Research
EGEE	Enabling Grids for E-sciencE
EGI_DIS	European Grid Initiative Design Study
EM	Environmental Monitoring
EO	Earth Observation
ESA	European Space Agency
ETICS	eInfrastructure for Testing, Integration and Configuration of Software
FAO	Food and Agriculture Organization
FCPPS	Fisheries Country Profiles Production System
FIE	Fisheries and Aquaculture Economics and Policy Division
FIES	Fisheries and Aquaculture Information and Statistics Service
FIM	Fisheries and Aquaculture Management Division.
FIRMS	Fisheries Resources Monitoring System
GENESI-DR	Ground European Network for Earth Science Interoperations - Digital Repositories
GMES	Global Monitoring for Environment and Security
Grid-QCM	Grid Quality Certification Model
GRL2020	Global Research Libraries 2020
ICES	International Council for the Exploration of the Seas
ICIS	Integrated Capture Information System
KCEW	Knowledge Exchange and Capacity Building Division
LEGN	FAO Legal department

NAFO	Northwest Atlantic Fisheries Organization
NEAFC	North East Atlantic Fisheries Commission
NeOn	Lifecycle of Networked Ontologies
OGF	Open Grid Forum
OGC	Open Geospatial Committee
OGSA	Open Grid Services Architecture
Planets	Preservation and Long-term Access through NETworked Services
SAPIR	Search in Audio-Visual Content Using Peer-to-Peer IR
VE	Virtual Environment
VO	Virtual Organization

TABLE OF CONTENT

SUMMARY

This deliverable reports on the collaborations with other FP6 & FP7 projects and R&D programmes established by D4Science from the beginning of the project to September 2008. These collaborations are of different nature, as they range from purely technical exchanges involving mutual exploitation of technologiesto the sharing of e-Infrastructure resources and to the joint organisation of networking and dissemination events. The deliverable presents these collaborations clustered into: (i) *Technological projects*, which either develop technologies that can be exploited by D4Science or that use the D4Science developed ones (e.g. EGEE III, Health-e-Child, ETICS-2, SAPIR, DRIVER II, NEON); (ii) *User Communities related projects,* which comprise projects and initiatives that have started collaborating with D4Science through the mediation of the scientific communities represented in the project (e.g. FIE, FIM, FIES, CGIAR, NAFO, CWP); (iii) *Other projects and initiatives,* which include EU projects and initiatives addressing aspects complementary to the technological one, like exploitation, outreach and sustainability (e.g. OGF-Europe, BELIEF-II, DL.org) and iv) *Standardization Working Groups,* which refers to on-going collaborations with standardization Working Groups (e.g. OGF Data Working Group, Dublin Core AgMES-Agricultural Metadata Element Set).

We expect that the number of collaborations will be increased in the next future as an effect of the availability of the production infrastructure. These new collaborations will be reported in an update version of this deliverable that will be published at M20.

EXECUTIVE SUMMARY

This document reports on the collaborations with other FP6 & FP7 projects and R&D programmes established by D4Science from the beginning of the project to September 2008. These collaborations are of vital importance for the D4Science project for three main reasons: First, D4Science concentrates on building a new complex Service-Oriented Architecture technology and an e-Infrastructure that addresses different levels of functionality. Resources, technologies and standards provided by other projects are essential for enriching the capabilities that the e-Infrastructure can offer to its users. Secondly, D4Science aims at deploying a networked, grid-based, and data-centric e-Infrastructure capable of supporting a wide variety of new concrete multidisciplinary scientific domains. Collaborations with projects that address user communities' needs enables D4Science to promote the developed e-Infrastructure, to evaluate its choices and receive feedback on how it can be exploited in other domains. Finally, the sustainability of the D4Science e-Infrastructure largely depends on the number of users and applications it will be able to attract. Collaboration with projects that offer D4Science opportunities for promoting its outcomes largely facilitates a wider outreach. The collaborations presented are grouped into the following clusters:

- *Technological projects.* This cluster includes European projects which either develop technologies that can be exploited by D4Science or that use the D4Science developed ones. Some of these collaborations were established during DILIGENT ("A Testbed Digital Library Infrastructure on Grid Enabled Technolgoy", FP6 Contract No.004260), the D4Science predecessor. These are now quite consolidated, with frequent exchanges of technologies and resources (EGEE III, Health-e-Child, ETICS-2, SAPIR). Others have been initiated more recently. These concern learning/analysis of the existing technology and shared plans for actual/potential adoption of it (DRIVER II, GENESI-DR, SHAMAN, CLARIN, CASPAR and NEON).

- *User Communities related projects*. This cluster groups projects and initiatives that have started collaborating with D4Science through the mediation of the scientific communities represented in the project, i.e. Environmental Monitoring, Fisheries Country Profile Production System and Integrated Capture Information System communities. Some of these projects and initiatives are managed by FAO Divisions (e.g. FIE, FIM, FIES, LEGN, KECW), while others are managed by international organizations affiliated with the scientific communities (e.g. CGIAR, NAFO, ICES, EuroStat, NEAFC, CWP, FIRMS). In many cases, these communities also span across different domains, like Marine and Ocean, Geophysics and Astrophysics.

- *Other projects and initiatives.* This cluster includes other EU projects and initiatives addressing aspects complementary to the technological one, like exploitation, outreach and sustainability. Co-operations with projects like OGF-Europe, BELIEF-II, the newly funded DL.org and the GRL2020 initiative largely facilitates D4Science networking and dissemination. D4Science exploits the framework offered by these projects to disseminate its outcomes, while these projects benefit from the contribution that D4Science brings to discussions on the topics addressed in the organized events. On the sustainability side, the recently established collaboration with EGI_DS is expected to be largely intensified in the future.

- *Standardization Working Groups*. This last cluster refers to on-going collaborations with standardization Working Groups (e.g. OGF Data Working Group, Dublin Core AgMES-Agricultural Metadata Element Set) that are relevant to the different aspects of interoperability addressed by the project. This type of collaboration is very important for D4Science which, since its conception, strongly relies on standards support because of the networked and distributed scenarios it is going to serve.

1 INTRODUCTION

1.1 Overview

Several collaboration exchanges have been established in the course of the D4Science project with other FP6 and FP7 EU projects and international R&D research programmes. These exchanges cover all the activities of the project ranging from technology, to the sharing of resources and to networking with other communities. This deliverable presents the collaborations established from the beginning of the project to September 2008. Foreach collaboration, it briefly introduces the project, describes the type of collaboration established and the plans for the future, if any.

1.2 Deliverable outline

The deliverable is organised according to the typology of the collaborating projects. Section 2 describes the collaborations with technological projects; Section 3 presents the collaborations with projects related to the D4Science user communities; Section 4 deals with projects specialised on supporting across-projects dissemination, sustainability, and exploitation; Section 5 focuses on the collaborations with standardization Working Groups; and finally, Section 6 concludes and lists plans for starting up effective co-operations with other projects.

2 TECHNOLOGICAL PROJECTS

This section presents the collaborations established with European projects which either develop technologies that can be exploited by D4Science or that use the D4Science developed ones. For each collaborating project, a brief overview of its objectives is given, followed by a description of the issues addressed in the co-operation, the exchanges made and the plans for future, if any. In describing the collaboration, a distinction is made between collaboration on technological aspects (Technology), on the Infrastructure as a service (Service), and on dissemination and networking activities (Networking).

2.1 DRIVER

Project description

DRIVER II [1], the sequel to IST-FP6 DRIVER (Digital Repository Infrastructure Vision for European Research) project, is an EU7FP "e-Infrastructure - Capacity Programme" co-funded project. It involves 12 partners from 10 European countries. It aims at extending geographically the European Repository Infrastructure created by its predecessor while putting it into production. Currently, this e-Infrastructure is composed by more than 140 archives coming from 20 EU Countries which contain approximately 700.000 documents. From the technological point of view, DRIVER is implemented following a Service Oriented Architecture paradigm. It offers rich functionality for structuring, maintaining and serving content, spanning from simple documents up to enhanced ones, i.e. complex entities with bindings to source information. A DRIVER application is now available on the production infrastructure. This application gives access to the entire information space available through the infrastructure and offers personalized information retrieval facilities. Other applications, activated on the same infrastructure, are under construction.

Collaboration

The fact that D4Science and Driver II share two major partners, namely CNR-ISTI and NKUA, makes the collaboration between the two projects quite effective. This collaboration covers different aspects:

Technology

Despite the fact that DRIVER II and D4Science build on different technological frameworks, the former one being a generic SOA infrastructure, the latter being based on OGSA principles, there are several areas where the two projects share technological interests. Both have to work around common technological issues and build concepts that will capture the new trends of the entire Knowledge Management domain upon which they act. Up to the recent past, the direct exchange of code artifacts has been difficult due to differences in the technological and implementation roadmaps followed by the two projects. However this gap is currently being gradually bridged with various moves from two sides, such as adoption of the JAVA platform also in the DRIVER project, and D4Science building bridges for gCube to other systems, etc.

Service

The nature of the resources available in the DRIVER and D4Science Infrastructure is different, yet the two infrastructures share common challenges. Most importantly, they have to identifying the procedures and policies that are required to regulate the registration and sharing of data resources in this new type of infrastructures. They also have to design and develop services for managing and monitoring the infrastructure. These are both areas where the two projects plan to strictly collaborate.

Networking

The two projects also collaborate in networking activities, participating in common events and mutually promoting each other's work. D4Science also closely follows the activity dedicated to the definition of as sustainable model based on the DRIVER Confederation.

The collaboration is quite tight and includes regular physical cross-team meetings internal to the CNR and NKUA (2 to 3 meetings per month) and more formal, cross-organizational ones, in a sparser schedule. Topics that are discussed in these meetings include Resource Management, Information Model, Content Management, Enhanced publications / Live documents / Complex Documents concepts, mechanisms such as Indexing and ResultSets etc., all of which mutually influence implementation.

Plans for the future

The short/mid-term plans for the collaboration focus especially on the technical aspects. D4Science, which started earlier than DRIVER, has acquired experience on specific e-Infrastructure management, that can be usefully transferred to the DRIVER e-Infrastructure. Moreover, D4Science has already faced issues related to the management of multitype data content, and related issues and solutions can be useful for the support of enhanced documents as defined in the framework of the DRIVER project. In the long term, the cross-project access to services is foreseen, allowing DRIVER infrastructures to access the vast grid resource pool and D4Science to provide the VREs with access to the large information space of the DRIVER repositories.

2.2 EGEE-III

Project description

EGEE-III [1] is an EU 7FP "e-Infrastructure - Capacity Programme" co-funded project. It is the third phase of the EGEE (Enabling Grids for E-sciencE) project and, in this phase it involves 70 partners. This project developed and is maintaining the largest multi-disciplinary grid infrastructure in the world, which brings together more than 120 organisations to make a reliable and scalable computing resource available to the European and global research community. At present, it consists of 250 sites in 48 countries and more than 68,000 CPUs available to some 8,000 users 24 hours a day, 7 days a week.

The current phase has two clear objectives that are essential for research infrastructures:

- to expand, optimize and simplify the use of the worldwide largest production Grid by continuous operation of the infrastructure, support for more user communities, and addition of further computational and data resources;
- to prepare the migration of the existing Grid from a project-based model to a sustainable federated infrastructure based on National Grid Initiatives.

Collaboration

The D4Science predecessor, DILIGENT, was the first project external to EGEE to exploit the EGEE middleware (gLite) and the EGEE Infrastructure. This achievement has been obtained thanks to the close collaboration between the two projects. This strict collaboration continues also at the present. It spans in different areas of work and involves different people.

Technology

The D4Science project collaborates with EGEE Joint Research Activities by directly exploiting the gLite middleware. In particular, the D4Science software interfaces with the gLite Workload Management System, the gLite Disk Pool Manager, and the gLite Virtual Organization Membership Service. This collaboration requires the interaction

between the development teams of both projects, usually via the EGEE issue tracking system.

During the reporting period a more specific collaboration has been initiated to understand and use the GridMap application. This application, developed as part of the EGEE project, supports the monitoring and visualization of the status of the infrastructure. A number of phone meetings have been organised with the GridMap developers in order to become familiar with the technology and understand how it could be used and extended to cover the D4Science monitoring needs.

Service

From the infrastructure perspective, D4Science actively contributes to the EGEE production infrastructure by providing four gLite sites and therefore extending the EGEE computing and storage resources. These sites provide almost all services of the gLite 3.1 release. The deployment and certification of these D4Science gLite sites is done in strict collaboration with different EGEE Regional Operations Centres (ROCs) and relies on the procedures and tools defined by EGEE for the operation of gLite sites.

On the other hand, the d4science.research-infrastructures.eu Virtual Organisation has been approved by the EGEE Service Activity, allowing D4Science users to access the resources of the EGEE infrastructure where the D4Science VO is recognized. This has been achieved via direct collaboration with the EGEE Service Activity and the EGEE VO Managers groups. D4Science is one of few cross-sites and cross-ROCs VOs. Setting it up has required several interactions. As a result of this activity a lot of feedback has been provided to the EGEE Service activity.

Networking

There is also a strong collaboration between the two projects with regard to dissemination. D4Science is regularly present at EGEE Conferences and EGEE User Forums. This participation is extremely useful to synchronize the activities of both projects and better plan the future activities. In particular, D4Science participated in the third EGEE User Forum (Clermont-Ferrand, February 2008) by presenting its work on metadata management and actively participating in the organisation of the session dedicated to the data management. The project also participated in the EGEE Conference (September 2008) by organising two sessions titled ”Scientific Data Infrastructure Ecosystem”. This session was attended by approximately 50 people, most of them being EGEE users or technology providers. As a result of this session a number of synergies with other projects was discussed.

Plans for the future

It is expected that the fruitful collaboration with the EGEE project will continue. Now that the technology in both projects is more stable, the collaboration is expected to primarily concern the Service and Networking activities. In particular, the project is already working to the preparation of session proposals for the upcoming "EGEE-User Forum/OGF25” that will be held in March 2009 in Catania (Italy).

2.3 GENESI-DR

Project Description

GENESI-DR (Ground European Network for Earth Science Interoperations - Digital Repositories) [3], is an EU 7FP “e-Infrastructure - Capacity Programme” co-funded project. It aims at establishing open Earth Science Digital Repository access for European and world-wide science users. In particular, it addresses the following objectives:

- To provide guaranteed, reliable, easy, effective, and operational access to a variety of data sources, and demonstrate how the same approach can be extended to provide access to all Earth Science data.
- To harmonise operations at key Earth Science data repositories limiting fragmentation of solutions. To demonstrate effective curation and prepare the frame for approaching long term preservation of Earth Science data.
- To validate the effective capabilities required to access distributed repositories for new communities, including education, and assess benefits and impacts.
- To integrate new scientific and technological derived paradigms into operational infrastructures in response to the latest Earth Science requirements.

GENESI-DR builds upon the existing, operational and focused Earth Observation (EO) European infrastructure and involves key Earth Science centres responsible for operational data acquisition, processing, archiving and distribution.

Collaboration

D4Science shares GENESIS-DR's objectives aiming to provide guaranteed, reliable, easy, effective, and operational access to a variety of Earth Science data sources, for further integration with other types of information (e.g. documentation, other digital objects) needed by the Earth Science and other communities to build dynamic Virtual Research Environments. In this sense, the collaboration between the two projects shall exploit the power of Web Services in general and Web Services Processing (WSP) features of interest for both Grid and Earth Science communities. Note that standards for WSP are being discussed by the OGF (Open Grid Forum) and OGC (Open Geospatial Committee).

Technology

The GENESI-DR infrastructure refers to the open European-wide federation of distributed Earth Science facilities, interconnecting in particular GENESI-DR participating sites and the participating Earth Science Digital Repositories, facilitating transparent access to data and associated GENESI-DR services.

Among available capabilities and ongoing developments that constitute the initial reference infrastructure of GENESI-DR there is the ESA EO Grid on Demand infrastructure[1] which integrates high-speed connectivity, distributed processing resources and large volumes of data to provide science and industrial partners with improved access to end products.

D4Science is continuing a close collaboration with the ESA team responsible for the EO Grid infrastructure as initiated with DILIGENT: current activities involve the direct submission of jobs to the ESA Grid infrastructure and the integration of results within the D4Science repository, thus fully exploiting the Grid paradigm behind both projects. D4Science is following a precise roadmap to exploit the foreseen GENESI-DR infrastructure once it becomes available through user services.

Service

The GENESI-DR services are provided to the community to better exploit Earth Science data covering functionalities such as: discovery, access to data, on-demand processing, storage of user results. A major objective of GENESI-DR is to integrate new scientific and technological derived paradigms in operational infrastructures in response to latest Earth Science requirements: one of the preliminary measures to fulfil this objective is to adapt the ESA Grid on-Demand for multiple digital repositories by supporting at least one application based on Grid-on-Demand access to multiple sites.

[1] http://eogrid.esrin.esa.int

Through services for the monitoring of land vegetation and ocean chlorophyll, D4Science currently advertises a direct interaction with the ESA EO Grid on Demand Portal which offers access to science-oriented Earth Observation Grid services and applications including access to a number of global geophysical ENVISAT products.

Once ESA EO Grid on Demand Portal user services are adapted to work on multiple digital repositories D4Science can provide the means for a dynamic integration of those repositories providing the communities with a broader view on data access and exploitation; on the other hand D4Science can offer a reliable operational environment for the storage, processing and manipulation of results delivered by those services.

Networking

On various occasions, detailed discussion for effective collaboration between the two projects was performed in meetings such as:

- Concertation Event organized by BELIEF II June2008
- EGEE User Forum (presentation to a session organized by D4SCIENCE was delivered by GENESI-DR), September 2008
- GRL2020 workshop, March 2008

Plans for the future

A possible plan for future collaboration discussed with the GENESI-DR partners is to identify how D4Science might programmatically exploit the GENESI-DR resources (data and products). This would make it possible to largely enrich the D4Science offer and, in particular, the EM and FARM VREs.

2.4 SHAMAN

Project Description

The SHAMAN project [4], with its 17 partner members, is an Integrated Project co-financed by the EU 7FP "Cultural heritage and technology enhanced learning" Unit. Its aim is to develop a next generation Grid-based infrastructure for long term digital preservation (DP), including tools for analysing, ingesting, managing, accessing and reusing information objects and data across libraries and archives. To achieve its goals, SHAMAN applies grid-based multivalent, linguistic, semantic, and peer-to-peer methods for DP within its service-based infrastructure. The core services of the SHAMAN framework are constructed by integrating Data Grid, Digital Library, Persistent Archive, Context Representation, Deep Linguistic Analysis, and corresponding Semantic Representation and Annotation technologies. Three prototypical applications will support trialling and validation of the results in scientific publishing, libraries, parliamentary archives, industrial design and engineering and, experimentally, scientific application domains.

Collaboration

At this stage, the collaboration has been informal and mainly focuses on technological aspects.

Technology

D4Science and SHAMAN are strongly linked by the broad goal of deploying service-oriented, application-level e-Infrastructures based on Grid principles and middleware technologies. As such, they share a functional focus on content/knowledge orientation and data, cycle, and interaction intensive processes. SHAMAN specializes functionally on preservation though it is cross-sector in scope. D4Science is more broadly supportive of content-oriented functionality but orientates its services towards scientific research collaborations. Naturally, the architectures overlap in functional components, primarily in

the core areas of data management and security. The collaboration between the two projects has been initiated on this aspect. SHAMAN, which is in an early stage of development, has expressed strong interest in the architectural principles which govern the D4Science infrastructure with a view to quickly bootstrapping the definition and implementation of its own architecture.

Until now, three informal technical meetings have been organised in Glasgow (March, April and June 2008). All of them have been focussed on discussing the approaches to architectural specifications.

Plans for the future

D4Science will monitor closely the evolution of SHAMAN with a view to supporting the project in its architectural choices and in building interoperability bridges across the two infrastructures so as to increase the pool of resources and specialise the suite of services which it offers to its current and future communities.

2.5 CLARIN

The CLARIN (Common Language Resources and Technology Infrastructure Network)project funded by the "ICT-Research Infrastructures" Unit [5]. It is a large-scale pan-European collaborative effort to create, coordinate and make language resources and technology available and readily usable for Language & SSH (Social Sciences & Humanities) researchers. It aims at offering scholars the tools to allow computer-aided language processing, addressing one or more of the multiple roles language plays (i.e. carrier of cultural content and knowledge, instrument of communication, component of identity and object of study) in the Humanities and Social Sciences. To achieve these challenging goals CLARIN will be built on a number of key technologies coming from the major initiatives advancing the eScience paradigm:

- It will include Data Grid technology to connect the repositories as being implemented in the DAM-LR pilot project [48] and web services the various centres participating in the project provide;
- It will build on ideas launched by the Digital Library community to create "Live Archives," and will further advance such initiatives;
- It will incorporate, and contribute to, Semantic Web technology to overcome the structural and semantic encoding problems;
- It will incorporate advanced multi-lingual language processing technology that supports cultural and linguistic integration.

Collaboration

The collaboration with CLARIN is in an early phase and concerns mainly the possible exploitation by CLARIN of e-Infrastructure specific technologies developed by D4Science in order to address the needs of the linguistics researchers community.

Technology

Many features strongly required by CLARIN, like "virtual" collection of resources from different repositories, workflow specification tools for processing this virtual collection with possibly a mix of home grown and remote service components, access control to resources, are already available in the D4Science e-Infrastructure. D4Science can thus

offer to CLARIN the solutions developed while CLARIN can offer an interesting context to validate them.

Plans for the future

A presentation of the D4Science Technology will be given at the CLARIN experts Workshop that will be organized by this project in February 2009. This meeting will clarify to what extent the collaboration can be intensified.

2.6 Health-e-Child

Project Description

Health-e-Child (An integrated platform for European Pediatrics based on a Grid-enabled network of leading clinical centres) [6] is an FP6-IST project co-funded by the ICT for Health Unit. It aims at developing an integrated healthcare platform for European pediatrics, providing seamless integration of traditional and emerging sources of biomedical information. The total project cost is M€ 16.7 and very important industrial and research entities in informatics, biometrics and healthcare are members of its consortium.

Collaboration

The common interest among D4Science and Health-e-Child is mainly on technological aspects.

Technology

The D4Science and Health-e-Child projects share the base technology for the information/data processing pipeline, both building on top of the Grid and the OGSA paradigm and the Globus WS-Core framework. Since the beginning of the projects the flow of technology from D4Science to Health-e-Child started with quite concrete artifacts and resulted in rich feedback flowing back from the Health-e-Child implementation teams, which significantly improved the quality and completeness of fundamental gCube components, such as the ResultSet and the search engine. Nowadays, D4Science is closely observing the evolution occurring within Health-e-Child so as to adopt the new significant artifacts developed there, which stem from the special requirements of the Bioinformatics and Medical community served by the project. In this direction it has proved extremely important that the two projects share a common implementation base, since this significantly reduces the costs of materializing the collaboration results.

NKUA, a common partner of Health-e-Child and D4Science, is the flow channel of information and technology among the two projects. Regular meetings occur nearly every week not only for exchanging information on each other's progress, problems and challenges but also for studying fields of common interest for future implementation.

Plans for the future

The most important part of the future implementation is the exchange of experiences and components, which in the long term could evolve into a common OGSA-based information processing engine.

2.7 ETICS 2

Project description

The ETICS 2 project [7], the ETICS project's second phase, is co-funded by the EU 7FP "e-Infrastructure - Capacity Programme". ETICS 2 started on 1st March 2008 and will last for 24 months involving 8 partners, including an international organization and 3 national research organizations. ETICS stands for "eInfrastructure for Testing, Integration and Configuration of Software". It provides a service to help software developers, managers and users to better manage complexity and improve the quality of their software. Using cutting edge Grid software and best practices, the ETICS service allows users to fully automate the way their software is built and tested. In other words, ETICS provides software professionals with an "out-of-the-box" build and test system, powered with a build and test product repository.

ETICS is multi-platform and open source. The client is designed to be simple to install. Results from daily, nightly and continuous builds and tests can be monitored via the web. Users can also browse and edit project data via a secured web application.

Collaboration

The collaboration with ETICS mainly concerns the technological area:

Technology

D4Science uses ETICS services as its build and integration platform. The current ETICS system is the result of two years of continuous refinement, where the collaboration with the DILIGENT (the predecessor of the D4Science project) team was a continuous source of feedback. D4Science has now fully standardized its software engineering process on it.

The D4Science integration and testing activity relies on the ETICS system for most of its tasks. This collaboration involves the different teams working in the preparation of the gCube releases. In particular:

- Developers rely on ETICS to execute local and remote builds of the component they are developing. ETICS is also useful for developers to share among them the different gCube components.
- The Build team exploits the ETICS system in the integration of the different software components provided by developers and builds a common gCube release. ETICS allows the execution of local and remote builds and the definition of different types of builds (continuous builds, integration builds, etc) by creating different configurations for the same component.
- The testing team uses ETICS for the execution of deployment tests and static code analyzers. The local or remote execution of these tests using ETICS facilitates and automatises the tasks of D4Science testers.

This strong collaboration is based on the work previously carried out by the DILIGENT and ETICS projects, and continues in D4Science with frequent discussions between the two projects where feedback is provided and new requirements discussed. ETICS-II also provides D4Science with efficient support whenever assistance is required.

In particular, D4Science participated to the ETICS session in EGEE'08 (September 2008) by providing feedback about D4Science usage of the system. In particular, the current usage of ETICS for building and testing gCube and gCore was explained. Possible future improvements in ETICS were presented and discussed with the ETICS team.

Plans for the future

ETICS 2 is currently setting up the implementation of the Grid Quality Certification Model (Grid-QCM) on its build and test framework. Grid-QCM has been developed within the ETICS activities in order to exploit distributed infrastructure providing an automatic way to evaluate, certify and compare software regardless of the processes used to develop it. The ETICS 2 staff is thinking of D4Science as a candidate to test and tune the implementation of the model through a trial certification issuing activity. D4Science shall

be tested against the Grid-QCM metrics getting back the results in order to improve the system, if it is necessary, or obtain the Grid-QCM (beta) certification.
Finally, ETICS 2 shall discuss the results with D4Science expert people to improve the model implementation.

2.8 SAPIR

SAPIR (Search in Audio-Visual Content Using Peer-to-Peer IR) [8] is an EU FP6 project co-funded by the "Networked Media" Unit. It aims at extending the power of web searches beyond centralized text and metadata to include distributed audio-visual content. It is conducted by 9 partners, three of which are relevant international industries. Implementing real content-based, audio-visual searches requires media-specific understanding and extremely high CPU utilization, which do not scale in today's centralized solutions. SAPIR aims at breaking this technological barrier by developing a large-scale, distributed P2P architecture that makes it possible to search audio-visual content using the query-by-example paradigm (e.g. a picture by a cell phone). To achieve this goal, a common framework for feature extraction from all media contents is being developed allowing similarity search and ranking along all supported media.

Collaboration

The collaboration with SAPIR mainly concerns the exploitation of the D4Science e-Infrastructure.

Service

In order to measure and validate the quality of the developed search engine, SAPIR has initially planned an experimentation with 100 Millions images gathered from Flikr. Extracting features from this large collection is very computational intensive, therefore the processing capacity needed to extract features from this large collection was exceeding the resources available to the project partners. In order to overcome this problem, SAPIR established a collaboration with DILIGENT and with its successor, D4Science, for the hosting of the feature extraction services on the existing infrastructure. By exploiting the D4Science infrastructure, 22 Millions of images were processed in the period August–September 2008. This processing has generated 67 Millions of objects that use approximately 2.7 TB of storage.

Plans for the future

The features extracted by SAPIR are not only useful to its objectives, but also to any other project that intends to experiment and validate indexes on such type of features. To make this possible, the SAPIR project has decided to render the set of the extracted features publicly available as a service to the entire research community. Having in mind this objective, the possibility of applying the extraction of features also to a larger image collection is being explored. D4Science has expressed its availabilty to host the necessary computations also in this SAPIR extension.

2.9 CASPAR

Project Description

CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) [9] - is an EU FP6 Integrated Project co-funded by the "Cultural heritage and technology enhanced learning" Unit . CASPAR's main objectives are as follows: Implement, extend, and validate the OAIS reference model [49] and design virtualisation services supporting long term digital resource preservation, despite changes in the underlying computing (hardware and software) and storage systems. The

project also aims to raise awareness about the critical importance of digital preservation among the relevant user-communities and to facilitate the emergence of a more diverse offer of systems and services for preservation of digital resources.

Collaboration

D4SCIENCE intends to continue with the cooperation activities started during the DILIGENT project. CASPAR is building a complete framework to allow the preservation not only of the data but also of the associated knowledge. As discussed in a meeting held between the two projects, the features of a distributed Virtual Research Environment could represent an important added value to improve the CASPAR system and in the future the hypothesis of an integration of the services released by D4Science will be taken into consideration by the CASPAR team.

Though CASPAR is going to end in 2009, important deliveries from that project are going to be provided already by the end of 2008: those achievements include both software releases and advances in the digital preservation research area.

A dedicated collaboration meeting was held early 2008 at ESA to explore potential collaboration across the two projects.

Future plans

D4SCIENCE will be tuned into CASPAR's progress of activity during the next months (via issuing closed meetings, public events or teleconferences) and accordingly plan for an integration work scenario where CASPAR software services can effectively exploit the infrastructure maintained by the D4Science team.

2.10 NeOn

Project Description

NeOn (Lifecycle of Networked Ontologies) [10] is a 4-year EU-FP6 project co-funded by the "Content and Knowledge" Unit. It involves 14 European partners - universities, national research centres, private companies and international organisations - which aims at creating the first ever service-oriented, open infrastructure, and associated methodology, to support the overall development life-cycle of a new generation of large scale, complex, semantic applications, with the overall goal of extending the state of the art to ensure that economically viable solutions can appear on the market. These solutions will enable efficient implementations of semantic applications, in open environments such as the Semantic Web, in support of the automation of Business to Business relationships, and also in company intranets. NeOn supports this overall strategic goal by:

- developing a generic NeOn reference architecture which aims to provide a standard, plug&play framework for integrating ontology life-cycle components;
- ensuring that the NeOn vision is concretely instantiated in a concrete implementation of the architecture, the NeOn ontology engineering toolkit, which provides the first instance of a new generation of ontology management tools;
- capturing key engineering processes into a NeOn methodology that provides the necessary framework to organize and manage the development of semantic applications à-la NeOn.

Collaboration

The collaboration with the NeOn project is facilitated by the involvement of FAO in this project. The NeOn project is a mixture of application development, advancement of new theoretical models, creation of methodological processes, and case study building. D4Science plans to use the ontologies developed by FAO in the NeON project within the

Fisheries VRE's. These ontologies can be used to structure, relate, index and locate data. Ontologies can be used as concept lists for annotating resources in VRE's with appropriate keywords. As these keywords are structured in ontologies, they can be used to improve search results by suggesting related resources or categories (this is exactly how Amazon's related items system functions). The ontologies can also function as browsing mechanisms, enabling users to see the resources that are annotated with the ontological concepts. Moving beyond mere annotation, ontologies can be used as the structural backbone for VRE information objects themselves, defining objects and their relationships to one another, thus greatly enriching the D4Science object model.

The FAO, D4Science and NeOn teams work very closely, therefore exchanges and updates on the new developments are frequent.

Future plans

Experiments in using ontologies as a way of overcoming problems of data siloing are ongoing in the D4Science project. Tools have been created that extract ontologies from database and XML schemas. These "machine" ontologies can then be mapped to domain ontologies, allowing the elegant harmonisation of different information systems. One aspect that the project plans to investigate is how such ontologies could be used within D4Science as a solution to data harmonisation.

3 USER COMMUNITIES RELATED PROJECTS

Through the collaboration with the Environmental Monitoring and the Fisheries and Aquaculture Resources Management communities exchanges with many other projects active in the specific communities' scenarios have also been established. In this section, the collaborations established, organized by scenario, are briefly presented.

3.1 Environmental Monitoring

Community Description

Environmental Monitoring community is represented in the D4Science project by the European Space Agency [11] and refers to a pan-European and international Earth Science scientific community. In particular, it includes experts and scientists in the field who systematically exploit user services and resources available from ESA (GEOPortal [13] G-POD [14], Earthnet [15]) for the most disparate applications of Earth observation domain, from land and ocean to atmosphere monitoring. At the present user services are offered by ESA to more than 5000 projects which benefit from world-wide free data access via the Principal Investigator Portal [16]. Furthermore, the ESA EoPortal [17] provides many thousands of scientists with access to information at various levels of complexity, including access to products, services, databases, and documentation. Such scientists belong to world-wide research centres, institutions, industry and universities.

Collaborations

In a long-term perspective the EM community shouldbroaden to major actors and key-players for the exploitation of data and innovative services for the Earth observation in an operational context, whose potential acme is represented by the GMES [18] initiative (Global Monitoring for Environment and Security), now renamed KOPERNIKUS. To this end the identified target audience comprise the GMES's catchment area [19], European Environmental Organizations (Eumetnet [20], EuroGOOS [21], Eurogeographics [22] and Eurogeosurveys [23], EEA [24], EMSA [25], EUMETSAT [26]), and national space agencies (e.g. the German [27], French [28] and Italian Space Agency [29]).

Specific opportunities to discuss the potential use of the D4Science platform have been presented by ESA (L. Fusco) in various specialised communities, such as:

- Marine and Ocean – represented by the SeaDataNet project [30] at the IMDIS 2008 conference in Athens
- Geophysics – at the EGU2008 (European Geophysical Union) [31] in Wien
- Astrophysics – a community whose data and infrastructure requirements are very close to the Earth Science Community, at European Southern Observatory–ESO conference in Munich

3.2 Fisheries Country Profiles Production System

Community description

The Fisheries Country Profiles Production System (FCPPS) community is a group of FAO-Fisheries services and national experts working to streamline the production of country profiles in order to cover more countries, update profiles more frequently, using more data sources, and create them in more languages. Although country profiles are one of the most popular FAO information products, they are cumbersome to assemble. They rely on many data sources and content authors. A VRE has been created in the D4Science e-Infrastructure to support the FCPPS community. This VRE combines

different data sources into a single environment, thus simplifying the report generation. Through this VRE National experts will upload their fisheries and stock inventories together with other authorial content. FAO time-series statistics will be accessible and used by FAO's Fisheries Resources service to produce assessments. A coordinator will combine these outputs together with other VRE data such as FAOLEX (regulations), FAOSTAT (statistics) and EIMS (publications) to produce final reports.

Collaborations

A number of collaborations have been established with projects/initiatives/working groups which cooperate with the FCPPS community. The most significant are briefly presented below.

a. *FIE – Fisheries and Aquaculture Economics and Policy Division*. This group controls the production of country profiles. Several meetings have been held with the division director Jean-François Pulvenis beginning in February 2008. He has given formal support to the FCPPS prototype, and if successful it will become the department's chosen vehicle for the production of these profiles. A conference to further elaborate stakeholder roles in the prototype will be held on October 18th 2008.
b. *FIM – Fisheries and Aquaculture Management Division*. This group produces assessments on the state of stocks and fisheries based on the available data. This information provides a vital snapshot of the relative health of a countries fisheries and fish stocks. They also have agreed to participate in the prototype by using it as part of their assessment process. A key stakeholder, Cassandra de Young has been defined and several meetings have been held with the responsible members of this group beginning in February 2008. They will also participate in the conference to further elaborate stakeholder roles on October 18th 2008.
c. *FIES – Fisheries and Aquaculture Information and Statistics Service*. This group collects, maintains and disseminates over 50 years of fisheries statistics on capture, production, fleets and commodities. They have agreed to participate in the prototype and add needed web services in order to ease access to their time-series data. The service chief Richard Grainger has given his formal support to the prototype. They will also participate in the conference to further elaborate stakeholder roles on October 18th 2008.
d. *LEGN – FAO Legal department*. This group maintains an important database of national and regional legal regulations pertaining to fisheries, forestries, agriculture and livestock. These regulations as well as associated port-state measures, a new initiative to combat illegal fishing will be ported into FCPPS. They have agreed to allow programmatic access to their data.
e. *KCEW – Knowledge Exchange and Capacity Building Division*. This group manages the Corporate document repository that will be ported to FCPPS. It contains numerous titles specific to countries or regions covered by the country profiles and constitutes an integral source of data used by editors in constructing reports. This group is working on corporate-level country profiles. Several meetings have taken place since March 2008 to define how data sharing can take place between the groups using D4Science technology.

3.3 Integrated Capture Information System

Community description

The Integrated Capture Information System (ICIS) community is a community of scientists working in the area of fisheries resource management. A specific VO has been created in the D4Science e-Infrastructure to include members and resources pertaining to this community. This VO gives dynamic access and/or static import of multiple fisheries statistical time-series data sets at varying levels of granularity and conforming

to different storage models. These are harmonized into a data set that conforms to a common model and a common set of metadata. From this VO, one or more Virtual Research Environments (VRE's) can be created; e.g. FAO-ICIS, a VRE that uses known species distributions (FAO) and extrapolated distributions (AquaMaps) from known species occurrence data (OBIS) to reallocate catch data between high-seas and coastal exclusive economic zones (EEZ's) in order to improve assessments that use coarse-grained data.

As for the ICIS community, D4Science-specific collaborations have been established with projects/initiatives/working groups which cooperate with the ICIS community. The following are the most significant ones:

a. *WorldFish Center [32]*. As a part of the Consultative Group on International Agricultural Research (CGIAR), it carries out research-for-development to improve small-scale fisheries and aquaculture. D4Science participants WorldFish and FAO are two important actors in fisheries sustainability that have different focuses and different data that combined can help fisheries scientists and managers make more informed decisions. Several meetings (January 2008, June 2008) have taken place between Senior WorldFish Centre researcher Nicolas Bailly and FAO-FIES representatives including Senior Information Officer Marc Taconet and Head Statistician Sachiko Tsuji to detail the scenarios, datasets, schedules and arrangements for cooperation. WorldFish was also present at the ICIS requirements gathering meeting held at FAO headquarters in Rome from June 10-13 2008.
b. *NAFO - Northwest Atlantic Fisheries Organization* [33]. This regional fisheries body (RFB) manages one of the most productive marine areas in the world and is recognized for the leading role they have played within their community in pushing the boundaries of new technology to fulfil their mandate. Holders of important fine-grained statistical data on fish captures and vessel movements, their data combined with coarser FAO data improves the ability of scientists to make stock assessments, model ecosystems and monitor biodiversity. There have been several contacts so far: in early April the NAFO data manager George Campanis spent several days at FAO headquarters meeting with FIES data managers to discuss how best to share data and reference data within the ICIS VO as well as looking towards the future of using fine-grained vessel monitoring system data. The executive secretary Johanne Fischer represented NAFO at the ICIS requirements gathering meeting held at FAO headquarters in Rome from June 10-13 2008 in which they reiterated their strong interest in being part of the project and agreed to send their data manager George Campanis in Autumn 2008 for a second meeting to clarify the modelling and design issues for data harmonization.
c. *ICES - International Council for the Exploration of the Seas [34]*. This Regional Fishery Bodies coordinates and promotes marine research in the North Atlantic. Providing times-series data and assessments for over a hundred commercially important species it is the oldest intergovernmental organisation in the world concerned with marine and fisheries science. They regularly gather, store, disseminate and analyze statistical fisheries data. The link between FAO and ICES provides an important conduit of basic information on the state of many commercially important species. ICES and FAO have been in regular contact since February 2008, receiving information on their data services from data manager Neil Holdsworth, and information on their assessment data from Advisory Programme Officer Mette Bertelsen who was also a participant in the ICIS requirements gathering meeting held at FAO headquarters in Rome from June 10-13 2008.
d. *EuroStat* - The Statistical Office of the European Communities. This group is the statistical arm of the European Commission. The storehouse for ICES data and EC member countries fisheries data, they are peripherally involved in the project as a data service provider. This involvement could be of great future benefit to FAO as EuroStat holds and disseminates a wide-ranging set of statistics coming from the

EU's 27 member states, while also acting as a standards and harmonization body. Contact has been made with Head of Fishery Statistics Section Franco Zampogna, and notably discussions are ongoing to define and use common standards and services such as SDMX (Statistical Data and Metadata Exchange) for time-series data in order to ease data harmonization in D4Science and other projects. This initiative will be presented internally at EuroStat during the Directors' Meeting secretariat of October 2008.

e. NEAFC - *North East Atlantic Fisheries Commission* [35]. It is a regional fisheries body formed to recommend measures to maintain the rational exploitation of fish stocks in the Atlantic and Arctic Oceans. An organization with a strong focus on Vessel Monitoring System data as a method to control fishing effort, they have expressed their strong interest in D4Science. FAO has agreed to keep them informed as to project progress, and to involve them more fully when ICIS moves in future steps towards VMS data.

f. *CWP - Coordinated Working Party*[36]. The Coordinating Working Party on Fishery Statistics (CWP) provides a mechanism to coordinate fishery statistical programmes of regional fishery bodies and other inter-governmental organizations with a remit for fishery statistics. Functional since 1960, the CWP's purpose is to continually review fishery statistics requirements for research, policy-making and management; agree on standard concepts, definitions, classifications and methodologies for the collection and collation of fishery statistics; make proposals for the coordination and streamlining of statistical activities among relevant intergovernmental organizations. In July 2008 D4Science and the ICIS case study were presented to the CWP during an intersessional meeting. It was agreed that the CWP would monitor progress, be informed by the ICIS group as to developments, and would give advice and make recommendations as the case study advances.

g. *FIRMS (Fisheries Resources Monitoring System)* - The primary aim of the Fishery Resources Monitoring System (FIRMS) is to provide access to a wide range of high-quality information on the global monitoring and management of fishery marine resources. The partnership draws together a unified group of international organizations and regional fishery bodies collaborating within a formal agreement to report and share information on fisheries resources. For effective fisheries information management, FIRMS also participates in the development and promotion of agreed standards. D4Science was presented to the 2nd FIRMS Technical Working Group meeting in Rome in April 2008, and again to the FIRMS Steering Committee meeting in Halifax in July 2008. Members expressed interest in the initiative and asked to be kept informed as to its progress.

4 OTHER PROJECTS AND INITIATIVES

This section describes the collaboration established with other EU projects addressing aspects complementary to the technological one, like sustainability, exploitation and outreach. Through the synergies established with these projects, D4Science aims at creating a better awareness on the project's tangible outcomes and on its proposed vision for scientific data e-Infrastructures of the future. EGI_DS

Project Description

The European Grid Initiative Design Study (EGI_DS) [40] is an EU 7FP "e-Infrastructure - Capacity Programme" co-funded project.. It represents an effort to establish a sustainable grid infrastructure in Europe. The main foundations of EGI are the National Grid Initiatives (NGI), which operate the grid infrastructures in each country. EGI_DS, involving 9 principal partner institutions, started on September 2007, and will last for 24 months.

The goal of EGI_DS is to evaluate use cases for the applicability of a coordinated effort, to identify processes and mechanisms for establishing EGI, to define the structure of a corresponding body, and ultimately to initiate the construction of the EGI organization. The EGI organization will ensure that Europe capitalises fully on its large investment in grid infrastructures, middleware development and applications. Under these lines, major objectives of the future EGI are:

- Ensure the long-term sustainability of the European e-infrastructure.
- Coordinate the integration and interaction between National Grid Infrastructures.
- Operate the European level of the production grid infrastructure for a wide range of scientific disciplines to link National Grid Infrastructures.
- Provide global services and support that complement and/or coordinate national services (Authentication, VO-support, security, etc.).
- Coordinate middleware development and standardisation to enhance the infrastructure by soliciting targeted developments from leading EU and National Grid middleware development projects.
- Link the European infrastructure with similar infrastructures elsewhere.
- Promote grid interface standards based on practical experience gained from grid operations and middleware integration activities, in consultation with relevant standards organizations.

Collaboration

EGI_DS is working closely with EGEE-III to learn from its experience in operating large scale international grid infrastructures. EGI_DS is relevant for D4Science, to clarify and facilitate the D4Science sustainability plans for its infrastructure and software. EGI_DS has recently prepared a blueprint for the future EGI organization (EGI.org). D4Science partners are following the evolution of this blueprint and also indirectly contributing to it.

Information about the D4Science project has been included in the EGI knowledge base (http://knowledge.eu-egi.eu/index.php/Main_Page) in the area providing a list of the main European e-Infrastructure projects.

D4Science partners have attended the first EGI_DS workshop held in Geneva, on June 30th 2008 (http://web.eu-egi.eu/events/workshops/geneva-2008/). This workshop concentrated on defining the structure of the EGI organization; the draft blueprint was discussed there. D4Science members established first contacts with EGI_DS activity leaders and discussed plans for the operation of the D4Science sites and software in relation to EGI. D4Science members have also taken active role in the EGI_DS workshop that took place in Istanbul in September 2008 (http://web.eu-

egi.eu/events/workshops/istanbul-2008/). This workshop concentrated on planning the transition to the future EGI model (EGI.org + NGIs). The objective of the Workshop was primarily to harmonise the effort within EGEE-III with the findings of the EGI Design Study and the policies endorsed by the National Grid Initiatives towards a sustainable European grid infrastructure. D4Science was invited to present its view on the transition to EGI in a panel with other major e-Infrastructure projects and large communities.

Plans for the future

The topics addressed by EGI_DS are of key importance for D4Science as well. In particular, D4Science is interested in following the progresses in the definition of the sustainability model proposed by EGI - and possibly establish synergies with it - in order to identify its own model. At the current stage, however, EGI_DS only focuses on more standard Grid infrastructures (i.e. those characterized by processing and storage resources). Instead, D4Science, as many other emerging e-Infrastructures that deal with more higher level resources (especially content), poses new challenges. Informal contacts with EGI_DS representatives have been established for collaborating on clarifying how the solution proposed by EGI might coexist with other models implemented by more higher level data e-Infrastructures.

4.1 OGF-Europe

Project Description

OGF-Europe (*Mobilising & Integrating Communities on Grids, Standards & Best Practices across Europe*)[41] is an EU 7FP "e-Infrastructure - Capacity Programme" funded project. It aims to play a key role in influencing the drive towards global standardization efforts and in bringing best practices back into the EU computing environment. It is interacting with diverse organizations across a wide range of domains to help create awareness of distributed computing and of the major benefits it delivers. In particular, it is aligned with OGF's global mission of pervasive Grid adoption through interoperable software standards. Key deliverables of the project include outreach seminars and workshops, adoption challenges and recommendation reports as well as community surveys, best practice reports and tutorials. OGF-Europe also co-ordinates an "Industry Experts" group to better understand how European enterprises are dealing with issues surrounding interoperations and standardisation and to engage them in the core work of OGF.

Collaboration

OGF-Europe gives D4Science a context where project partners can be informed on the evolution of emerging OGF standards. In this context, D4Science partners can also actively participate in the fine-tuning of these standards by reporting the feedback collected through the experience carried out within the project.
OGF-Europe provides D4Science also with several opportunities for networking with other projects and for performing dissemination. In particular, the participation in the Community Outreach Seminars is particularly meaningful for D4Science. These seminars aim to actively gather the people of OGF, OGF-Europe and user groups to explore key technology trends and build a consensus around key requirements.

D4Science representatives participated in the first Community Outreach seminar "Digital Repositories Interoperability Using Grid Technologies" in June 2008 (Barcelona) by giving two presentations. The collaboration between D4Science and the OGF Data Working Group (described in Section 5) was largely stimulated by this Workshop.

Plans for the future

D4Science will continue exploiting the support to networking offered by OGF-Europe. In particular, D4Science representatives have been invited to present the project in the Repository Curation Service Environments (RECURSE) Workshop on 1 December 2008, Edinburg, jointly supported by OGF-Europe and DReSNET. This workshop will focus on highlighting application environments where e-Science applications have much to gain, coupled with the benefits of open standards.

The organization of joint D4Science-OGF-Europe tutorials is also being explored.

4.2 BELIEF II

Project Description

The BELIEF-II (Bringing Europe's eLectronic Infrastructures to Expanding Frontiers - Phase II) [42] is an EU FP7 "e-Infrastructure - Capacity Programme" co-funded Coordination Action whose strategic objective is to promote the efficient and effective communication of results, networking and knowledge among EU e-Infrastructure projects and their users. In particular, it currently: a) provides a Digital Library (DL) which offers a uniform access to the material produced by these projects (e.g. deliverables, presentations, dissemination material); b) organizes a number of events dedicated to stimulate networking among these projects; and c) offers a suite of Joint Communication Products, such as the e-Infrastructure News Magazine Zero-In, can be used by other project to disseminate their outcomes.

Collaboration

Collaboration between D4Science and BELIEF II spans over all the three areas addressed by the project. In particular, D4Science exploits the BELIEF DL as a third-party service which not only disseminates the project's documents but also manages and administers them by providing an appropriate repository service. Through ad-hoc Web submission forms D4Science registered users (authentication and authorization are managed by the BELIEF DL on behalf of D4Science) can upload project documents into the BELIEF repository. These documents become then visible through the BELIEF DL and through the D4Science portal. This feature is enabled by a "plug-and-play" search capability embedded in the D4Science portal which allows D4Science portal users to transparently search for material of interest within the D4Science collection in the BELIEF DL.

D4Science also participated in the events organized by the BELIEF II project. In particular, in the reporting period it participated in the 5th e-Infrastructure Concertation Meeting "e-Infrastructures as Standardisation Drivers", in June 2008, Barcelona (Spain). This event was dedicated to reinforce and extend the cooperation between the major standard bodies and EU-funded projects for the effective development and wider use of EU e-Infrastructure and to intensify and optimise human interaction and synergies among projects in order to facilitate production, deployment and use of standards. These topics were particularly appreciated by D4Science since building on standards is one of the major D4Science goals.

Plans for future

D4Science plans to further exploit the dissemination and networking opportunities developed by BELIEF-II. In particular, the project will participate, together with other projects, in the 6th e-Infrastructure Concertation Meeting, on November 2008, in Lyon (France), contributing to the discussion about "Sustainability for e-Infrastructures". It also plans to disseminate its results by sending contributions to the Zero-in Magazine

and to the next edition of the "e-Infrastructure Success Stories" publication (both published by the BELIEF II project).

4.3 DL.org

Project Description

DL.org (Coordination Action on Digital Library Interoperability, Best Practices, and Modelling Foundations) is an EU FP7 Co-ordination Action co-funded by the "Cultural heritage and technology enhanced learning" Unit that will start in December 2008. It aims at creating a framework where key representatives from major initiatives and on-going digital library and repository infrastructure related projects may collaborate, discuss experiences, exchange expertise, work on interoperability of their solutions, promote shared standards, and provide the research community with a deeper understanding of key issues and new directions. The ultimate objective of this project is to integrate the many DL efforts that are currently on-going, each one acting in isolation and adopting ad-hoc solutions and methodologies. By operating such integration the project is expected to facilitate the rapid advance of research and the development of new techniques thus promoting the creation of a European Information Space. The main mechanism that will be used by the project for achieving its objective is the formation of six Thematic Working Groups (WGs), each one of which will target one of the six fundamental concepts of any information system: Content, User, Functionality, Quality, Policy, and Architecture. Each Working Group will consist of a small number of experts in the corresponding concept that will be drawn from prominent projects and initiatives currently running in Europe and around the globe.

Collaboration

D4Science is one of the projects that have agreed to participate in the DL.org working groups. It shares with DL.org a strong interest in all the research topics related to interoperability. By interacting with other projects D4Science can present its solutions and learn about solutions implemented by others. It is expected that from this collaboration common agreements can be established.

Plans for the future

D4Science will appoint one representative for each Thematic Working Group. Each representative will participate in the definition of the WGs charter, in which the specific interoperability issues addressed by the WG will be indicated, as well as in its activities. D4Science will also participate in other events organised by DL.org, like interoperability workshops, and will contribute to the definition of best practices in the area.

4.4 GRL2020

Description

GRL2020 (Global Research Libraries 2020) [43] is a series of workshops dedicated to paving the way for a Collaborative Global Research Environment for the future. These workshops aggregate world's most highly regarded experts in the field of Digital Libraries and other fields related to the management and access of scientific content.

The first workshop of the series was held in Washington State (USA) in September 2007; the second one in Pisa (Italy) in March 2008, and the third one will be in Taiwan in February 2009. Until now these workshops have been partially sustained by Microsoft Research. However, it is expected that this important series of workshop will continue also in the future sustained by International Funding Agencies.

Collaboration

The D4Science project objectives match very closely the ones addressed by GRL2020. D4Science is actually a prototypical realization of a Global Digital Library as defined by the participants in these workshops. D4Science partners are actively involved in the GRL2020 community. The project has been presented at the past workshops and interesting plans for collaboration between D4Science and GRL2020 workshop participants have been established. A report on the results of the last workshop is currently in preparation. On the one hand, this report includes the D4Science vision and on the other, it provides a useful source of inspiration for further extensions of the D4Science objectives.

Plans for the future

D4Science partners will be among the programme committee members for the next GRL2020 workshop and will participate in the definition of the topics addressed by this workshop.

5 STANDARDIZATION WORKING GROUPS

This section presents the standardization Working Groups in which D4Science partners are directly involved. These working groups are briefly introduced, then the collaborations established and the activities performed are described.

5.1 OGF Data WG

Description

OGF Data is a working group of The Open Grid Forum (OGF) [37]. The OGF community consists of thousands of individuals in industry and research, representing over 400 organizations in more than 50 countries. Together they work to accelerate adoption of grid computing worldwide, believing that grids will lead to new discoveries, new opportunities, and better business practices.

The scope of the OGF Data working group (WG) is to produce an overall data architecture for OGSA (Open Grid Services Architecture). It aims at identifying message patterns and interfaces that form part of that architecture. Where possible, it plans to use existing specifications to form appropriate parts of this architecture, liaising with the groups defining those specifications to encourage them to fit into the overall OGSA picture. Where existing specifications are not available, the WG will encourage the creation of new groups to fill the gaps.

Collaboration

D4Science is naturally participating to the OGF community, being strongly interested in driving the rapid evolution and adoption of applied distributed computing, critical to develop new, innovative and scalable applications and infrastructures. In such a context, D4Science people attended OGF23, one of the largest and most important events in the Grid area held this year, featuring an extensive community program for those interested in learning about Grid and distributed computing technologies. In this event, held in Barcelona (2-6 June 2008), the D4Science Project Director and Technical Director participated in the "Digital Repositories - Interoperability Using Grid Technologies" session by presenting the D4Science vision and outcomes. During this session, a decision was taken to initiate a collaboration between some of the participants (including D4Science members) and the members of the OGF Data WG. This collaboration will be dedicated to specifically address how federated repositories should relate to federated storage.

Plans for the future

D4Science intends to follow the evolution of the OGF standards and, in particular, the activities of the Data WG. A meeting of the members involved in data interoperability and integration will be held in a near future, most likely at the OGF25/EGEE User Forum (Catania, Italy) in March 2009.

5.2 Dublin Core

Description

The Dublin Core Metadata Initiative (DCMI) [38] is an organization dedicated to promoting the widespread adoption of interoperable metadata standards and developing specialized metadata vocabularies for describing resources that enable more intelligent

information discovery systems. The Dublin Core Metadata Initiative provides simple standards to facilitate the finding, sharing and management of information. DCMI does this by: i) developing and maintaining international standards for describing resources; ii) supporting a worldwide community of users and developers, and, iii) promoting widespread use of Dublin Core (DC) solutions.

The development and maintenance of a core set of metadata terms (the DCMI Metadata Terms) continues to be one of the main activities of DCMI. In addition, DCMI is developing guidelines and procedures to help implementers define and describe their usage of Dublin Core metadata in the form of Application Profiles.

Collaboration

The participation of D4Science in the discussions and decisions of the Dublin Core is vital to assure interoperability between the systems and networks powered by D4Science and the broad Open Access community using DC. DC Application Profiles are main instruments to ensure interoperability between web systems. D4Science will not only use them, but also influence them, when the need arises. D4Science participates in the DC initiative through FAO. FAO has participated in every major Dublin Core conference since 2001. It has developed an extension to Dublin Core (an application profile) for the needs of the agricultural community. Known as the Agricultural Metadata Element Set (AgMES), it uses all of the Dublin Core element set and adds agriculture specific extensions. Johannes Keizer (FAO) is a member of the Dublin Core advisory board.

Plans for the future

FAO plans to continue its collaboration and strengthen its ties with the DC Community. Johannes Keizer has recently taken on the role of chairing the DC-Knowledge Management Community. In this role he will certainly be able to favour the exchange between the DC community and D4Science.

6 CONCLUSIONS

This deliverable has presented the major cooperation links that D4Science has established with EU and International projects and initiatives until September 2008. In particular, it has described the types of on-going collaborations and plans for the immediate future.

Yet, these collaborations are not the only ones that have been identified since the beginning of the project. Among the collaborations that are under discussion, particularly interesting are those with projects addressing application of the D4Science e-Infrastructure in other domains. The biodiversity domain is certainly one of the most promising ones. Exchanges have been established with the *Biodiversity Heritage Library Project* (BHL) [44]. This project strives to establish a major corpus of digitized publications on the Web drawn from the historical biodiversity literature. Ten major natural history museum libraries, botanical libraries, and research institutions have joined to form this library, whose content will be made available through a global "biodiversity commons". Currently, this project is selecting an appropriate technology for realizing this global digital library. Currently exchanges between BHL and D4Science are on-going for establishing synergies between the two projects both in terms of technology and available content.

Another important project related to this domain is AquaMaps [45]. AquaMaps has developed a service for computer-generated predictions of natural occurrence of marine species, based on the environmental tolerance of a given species with respect to depth, salinity, temperature, primary productivity, and its association with sea ice or coastal areas. In order to generate better quality maps, AquaMaps needs a richer set of data resources and larger computational capabilities. A possible collaboration with the AquaMaps project is under discussion for enabling the maps generation on the D4Science e-Infrastructure.

The International Polar Year is a large scientific programme focused on the Arctic and the Antarctic from March 2007 to March 2009. Many are the projects that are collecting observation data in the framework of this programme for serving a multitude of scientists from over 60 nations examining a wide range of physical, biological and social research topics. The technical solution adopted by D4Science has been illustrated to leading representatives of this programme. These have expressed their interest in this solution. A possibly more effective collaboration is under evaluation.

On the more technical side, preservation is an area where interesting co-operations might be established. For the time being, D4Science is a member of the community set up by the DPE (Digital Preservation Europe) project [46]. This project fosters collaboration and synergies between many existing national and international initiatives across Europe to secure effective preservation of digital materials. DPE has recently established a collaboration agreement with the PLANETS [47] and CASPAR projects which focus on complementary preservation issues. Through a collaboration with DPE D4Science might also better reach these other projects.

The ones mentioned above are only some of the potential collaborations that have been discussed. The next version of this deliverable will report which one of these will have evolved in concrete and effective collaborations.

REFERENCES

[1] DRIVER Project. Web site: http://www.driver-repository.eu/
[2] EGEE-III Project. Web site: http://www.eu-egee.org/
[3] GENESI-DR Project. Web site: http://www.genesi-dr.eu/
[4] SHAMAN Project. Web site: http://shaman-ip.eu/
[5] CLARIN Project. Web site: http://www.clarin.eu/language-resources-and-technology
[6] Health-e-Child Project. Web site: http://www.health-e-child.org/
[7] ETICS 2 Project. Web site: http://www.etics.web.cern.ch/
[8] SAPIR Project. Web site: http://www.sapir.eu/
[9] CASPAR Project. Web site: http://www.casparpreserves.eu/
[10] NEON Project. Web site: http://www.neon-project.org/
[11] European Space Agency. Web site: http://www.esa.int/esaCP/index.html
[12] http://eogrid.esrin.esa.int
[13] GEOPortal. http://www.geoportal.org/web/guest/geo_home
[14] G-POD. Web site: http://eogrid.esrin.esa.int/
[15] Earthnet. Web site: http://earth.esa.int/
[16] ESA EO Principal Investigator Portal. http://eopi.esa.int/
[17] ESA EoPortal. http://www.eoportal.org/
[18] KOPERNIKUS. Web site: http://www.esa.int/esaLP/LPgmes.html
[19] GMES. Web site: http://www.gmes.info/160.0.html
[20] Eumetnet. Eb site: http://www.eumetnet.eu.org/
[21] EuroGOOS. Web site: http://www.eurogoos.org/
[22] Eurogeographics. Web site: http://www.eurogeographics.org/eng/00_home.asp
[23] Eurogeosurveys. Web site: http://www.eurogeosurveys.org/
[24] European Environment Agency. Web site: http://www.eea.europa.eu/
[25] European Maritime Safety Agency. Web site: http://www.emsa.eu.int/
[26] EUMETSAT. Web site: http://www.eumetsat.int/Home/index.htm
[27] German Aerospace Center. Web site: http://www.dlr.de/en/
[28] Cèntre National Ètudes Spatiales. Web site: http://www.cnes.fr/
[29] Agenzia Spaziale Italiana. Web site: http://www.asi.it/SiteIT/Default.aspx
[30] SeaDataNet project. Web site: http://www.seadatanet.org/
[31] EGU2008. Web site: http://meetings.copernicus.org/egu2008/
[32] WorldFish Center. Web site: http://www.worldfishcenter.org/
[33] Northwest Atlantic Fisheries Organization. Web site: http://www.nafo.int/
[34] Intern. Council for the Exploration of the Seas. Web site: http://www.ices.dk/
[35] North East Atlantic Fisheries Commission. Web site: http://www.neafc.org/
[36] Coordinating Working Party on Fishery Statistics. Web site: http://www.fao.org/fishery/cwp/meetings/en
[37] Open Grid Forum, Web site: http://www.ogf.org/
[38] Dublin Core Meatdata Initiative. Web site: http://dublincore.org
[39] European Committee for Standardization. Web site: http://www.cen.eu/
[40] EGI-DS Project. Web site: http://web.eu-egi.eu/
[41] OGF-Europe Project. Web site: http://www.ogf.org/ogfeurope.php

[42] BELIEF II Project. Web site: www.beliefproject.org/

[43] GRL2020. Web site: http://www.grl2020.net/

[44] Biodiversity Heritage Library Project. Web site: http://www.biodiversitylibrary.org/About.aspx

[45] AQUAMAPS Project. Web site: http://www.aquamaps.org/

[46] DigitalPreservationEurope Project. Web site: http://www.digitalpreservationeurope.eu/

[47] PLANETS Project. Web site: http://www.planets-project.eu/

[48] DAM-LR Project. Web site: http://www.mpi.nl/DAM-LR/

[49] Reference Model for an Open Archival Information System (OAIS) Blue Book. January 2002. http://public.ccsds.org/publications/archive/650x0b1.pdf