Diligent

A **DI**gital **L**ibrary **I**nfrastructure on **G**rid **EN**abled **T**echnology

Deliverable No 3.1.1:

"EGEE-DILIGENT Interaction Report"

July 2006

Document Information

Project	
Project Title:	DILIGENT, A **DI**gital **L**ibrary **I**nfrastructure on **G**rid **EN**abled **T**echnology
Project Start:	1 Sep 2004
Call/Instrument:	FP6-2003-IST-2/IP
Contract Number:	004260
Document	
Deliverable number:	D3.1.1
Deliverable title:	EGEE-DILIGENT Interaction Report
Contractual Date of Delivery:	Report 1: Month 8, Report 2: Month 22
Actual Date of Delivery:	Report 1: 15 June 2005, Report 2: 6 July 2006
Editor(s):	Roberta Faggian Marque, Pedro Andradde
Author(s):	Report 1: Roberta Faggian Marque, Pedro Andrade, Donatella Castelli, Pasquale Pagano, Henri Avancini, Bhaskar Mehta, Claudia Niederee, Adelheit Stein Report 2: Roberta Faggian Marque, Pedro Andrade, Donatella Castelli, Pasquale Pagano, Leonardo Candela, Bhaskar Mehta, Adelheit Stein
Participant(s):	CERN, CNR-ISTI, FhG/IPSI
Workpackage:	WP3.1
Workpackage title:	EGEE-DILIGENT Interaction
Workpackage leader:	CERN
Workpackage participants:	CERN, CNR-ISTI, FhG/IPSI
Est. Person-months for the writing of this report:	3.3
Distribution:	Public
Nature:	Report
Version/Revision:	1.2
Draft/Final	Final
Total number of pages: (including cover)	98
File name:	D3.1.1_M22_v1.2.doc
Key words:	*Digital libraries, EGEE, gLite, inter-project coordination.*

Document Log

Issue	Date	Comment	Author/Partner
M8	15/06/2005	Report 1 including information related to the period M0-M8.	WP3.1 participants
M22	23/06/2006	Report 2 including information related to the period M9-M22. Major changes: 1. Definition of a general structure including information related to the three periods 2. Update of introductory sections 3. Update of Chapter 1 with new interactions 4. First period report included under Chapter 2 5. Complete writing of Chapter 3	WP3.1 participants

Disclaimer

This document contains description of the DILIGENT 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.

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 DILIGENT 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 25 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/)

DILIGENT is a project partially funded by the European Union

Table of Contents

Figures

Tables

Summary

This deliverable describes the activities performed in the context of WP3.1.

WP3.1 objectives are:

- to coordinate the cooperation and the interactions between DILIGENT and the EGEE activities to enable the integration, exploitation and enhancement of EGEE achievements through DILIGENT objectives;

- to analyze the status of the available Grid infrastructure compared to the specific requirements of the digital libraries (DLs) community. Establish the basis for a profitable experimentation of Grid technologies in the knowledge management domain.

The purpose of this document is to describe how the interaction between the DILIGENT and EGEE projects takes place, as well as provide a first feedback to EGEE and the digital libraries community. This deliverable is an ongoing report and is progressively expanded as further experience is gained in DILIGENT.

More precisely, three versions of this report will be produced during the lifetime of the project: report 1 at project month 8, report 2 at project month 22 and report 3 at project month 36. These three periods are described in the document in chronological order in the three chapters: "Hands-on Technology", which covers the activities from month 1 to month 8, "Design and Experimentation", covering from month 9 to month 22, and "Built of the Prototype", from month 23 to month 36.
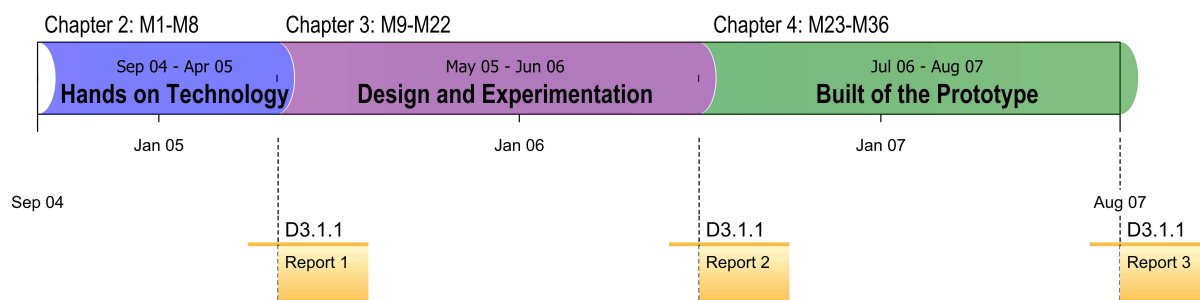
Figure 1: D3.1.1 on-going editing process.

Executive Summary

DILIGENT has chosen the EGEE project as the main middleware and infrastructure provider, the base for its future Grid developments. DILIGENT plans to adopt, experiment and extend EGEE achievements in the interest of the digital libraries community and users.

WP3.1 objectives are:

- to coordinate the cooperation and the interactions between DILIGENT and the EGEE activities to enable the integration, exploitation and enhancement of EGEE achievements through DILIGENT objectives;
- to analyze the status of the available Grid infrastructure compared to the specific requirements of the DLs community. Establish the basis for a profitable experimentation of Grid technologies in the knowledge management domain.

In order to achieve these objectives a close collaboration has been established between DILIGENT partners and EGEE representatives from the very beginning of the project through coordination meetings, technical discussions, training and dissemination events.

The purpose of this document is to describe how the interaction between the DILIGENT and EGEE projects takes place, how this interaction eases the collaboration between the two interrelated projects, how it allows to keep DILIGENT plans aligned with EGEE progress and how it permits to provide feedback to EGEE on various activities. This document is also intended to collect and provide feedback to EGEE and the DLs community on the activities performed so far bridging Grid and digital libraries domains. Finally this document aims to provide some useful references to other EGEE related projects, based on the current experience of DILIGENT.

Coordinating the relations between the DILIGENT and the EGEE projects is a complex activity since it requires actions at various levels and different areas. The tables reported in the first chapter of this report will confirm this fact by giving an overview of the past interactions. The editing of this document involves the collection of information which describes events and report outcome of experiences performed in the context of various DILIGENT workpackages and considered as relevant for the purpose of this deliverable.

This document is an ongoing report. Three versions of this report are produced during the lifetime of the project: report 1 at project month 8, report 2 at project month 22 and report 3 at project month 36. These three periods are described in the document in chronological order in the chapters Hands on Technology, from month 1 to month8, Design and Experimentation, from month 9 to month 22, and Built of the Prototype, from month 23 to month 36.

The EGEE activities are organised in Networking (management, dissemination, training, international cooperation, etc.), Services (operations, network resource provision, etc.) and Joint Research (middleware engineering, quality assurance, security, etc.). The structure of this document reflects this organisation in chapters treating these three areas of activity.

The first Chapter describes the main motivations and the general strategy adopted in establishing a fruitful relationship between DILIGENT and EGEE. Two summary tables provide a chronological overview of the interactions established between the two projects, further expanded later on in the document.

The following Chapters (2 and 3) present in detail the status of these activities, which are organised here according to the three main areas:

- Middleware, therefore everything in relation to the exploitation of gLite functionalities in DILIGENT;
- Networking, involving mainly the dissemination and training activities, but also including applications support and co-operations;
- Operations as related to the operational aspects of the infrastructure.

Interaction for each of these areas is necessitated for the following reasons:

- to be informed about recent developments and future plans, to get early access to documentation and software, to exchange technical views, to provide feedback and requirements on middleware, infrastructure and policies;
- to be informed about and involved in EGEE events, to organise joint events, to get early access to training material, to provide feedback as EGEE users;
- to be informed about EGEE production service policies, difficulties related to deployment and maintenance of the infrastructure.

Anther section in these chapters is intended to provide information related to the feedback DILIGENT provides to EGEE and the digital library community. It describes mechanisms for providing feedback, requirements, future opportunities for interactions, open issues and difficulties.

The last section of these chapters summarises the conclusions and highlights the fact that the two projects have taken mutual advantage from a deeper substantial technical collaboration improved over time.

1   DILIGENT-EGEE INTERACTION ACTIVITIES

This chapter describes the general strategy adopted in establishing a fruitful relationship between the DILIGENT and the EGEE projects. Two summary tables provide a chronological overview of the interactions made between the two projects and motivations for them. These tables list a number of events and summarise the activities performed by the DILIGENT project in the context of various workpackages.

1.1   Purpose

The DILIGENT-EGEE interaction activity has been established at different levels (web interfaces, mailing lists, personal contacts, etc.), on different areas (middleware, networking, services) and for different purposes. As the DILIGENT project advances in its activities, the motivations for interaction with EGEE evolve and can be satisfied in various ways. They are:

- **Learn** – accessing dissemination, technical material and software through the EGEE portal; attending EGEE training events, conferences and open meetings (all activity meetings, workshops, concertation events, etc.); organising joint trainings/workshops with EGEE representatives; subscribing to EGEE open technical mailing lists; studying the internal organisation of EGEE activities and infrastructures as a reference model; looking for opportunities for technical exchanges;

- **Experiment** – testing local installations of middleware components; having early access to a prototype testbed for a restricted number of users;

- **Provide Feedback** – submitting DILIGENT middleware requirements to the EGEE requirements database; establishing contacts and collaboration with the EGEE UIG (User Information Group) and NA4 representatives; documenting DILIGENT experience in this deliverable;

- **Deploy** – studying the installation guides; exchanging with EGEE technical representatives;

- **Implement** – exchanges with EGEE technical teams are foreseen to better design and implement the DILIGENT services;

- **Disseminate/Demonstrate** – presenting DILIGENT plans; demonstration of results could be foreseen for the future;

- **Align** – another important aspect is the necessity to keep DILIGENT plans aligned with EGEE (with respect to availability of the software, functionality, plans for future developments, policies, support, etc.) as well as to inform EGEE about DILIGENT priorities and requirements that could be taken into consideration together with the feedback of the other user communities.

A chronological view of DILIGENT-EGEE interaction activity is given in the next section.

The overall strategy of interaction between EGEE and DILIGENT consists of establishing contacts with representatives of EGEE activities and exploiting the official channels already put in place by the EGEE project for dissemination and user support. At the same time an attentive study of the project organisation, documentation and functioning is also necessary. Finally, whenever possible, establishing direct exchanges with EGEE representatives to improve inter-project collaborations (i.e. organisation of joint events) and creating opportunities for exchanges especially at the technical level. This last approach is needed particularly in the case of DILIGENT, being an early adopter of EGEE gLite Grid middleware. Considering the obvious timing necessary for a large collaboration like EGEE to

collect, organise and disseminate the information related to the most recent release of its middleware (which was not yet available when the DILIGENT project started), a close interaction between the two projects demonstrated to be extremely helpful to minimise the project's risks, especially in the design phase of DILIGENT, when the DILIGENT solution is conceived.

1.2   Overview of Interactions

Interactions between the DILIGENT and EGEE projects/representatives are listed in this section. These tables list a number of events and summarise the activities performed by the DILIGENT project in the context of various workpackages. Further details will be provided in the following chapters while describing the various activities performed per each area. This overview should give to the reader a global chronological view of the events, as well as an understanding of the motivation behind each of them, as described in the previous section.

First period: Month 0 to Month 8, Interactions 1 to 26

Second period: Month 9 to Month 22, Interaction 27 to 58

Third period: Month 23 to Month 36

Hands-on Technology (M0-M8)		
April 2004		
#1	**1st EGEE Conference**	Cork, Ireland
	DILIGENT partners attended the EGEE conference. DILIGENT project was presented at the plenary session.	
September 2004		
#2	**DILIGENT kick-off meeting**	Pisa, Italy
	EGEE JRA1 leader and Project Director attended the meeting and presented the status of gLite development.	
#3	**Joint technical meeting n.1**	Pisa, Italy
	Discussion on specific requirements of DILIGENT application with the EGEE JRA1 activity leader.	
#4	**EGEE All-activity meeting**	Geneva, Switzerland
	DILIGENT partners attended the EGEE all activity meeting to learn about status and planning of all activities.	
October 2004		
#5	**Joint technical meeting n.2**	Pisa, Italy
	Discussion of the issues related to the design of the DILIGENT Collective Layer Services with EGEE JRA1 leader and the coordinator of the EGEE American Federation.	
#6	**Subscription to EGEE user requirements database**	
	One DILIGENT representative has been nominated as responsible for submission of DILIGENT requirements to EGEE and given access to the system.	
#7	**ARDA workshop**	Geneva, Switzerland

	CERN partners attended the ARDA workshop.		
November 2004			
#8	**Joint technical meeting n.3**		Rome, Italy
	Discussion on DILIGENT technical vision with the coordinator of the EGEE American Federation.		
#9	**Joint technical meeting n.3**		Rome, Italy
	Brainstorming on DILIGENT Index and Search Service with the coordinator of the EGEE American Federation.		
#10	**2nd EGEE Conference and Concertation meeting**		The Hague, Netherlands
	DILIGENT partners attended the EGEE conference and concertation event. DILIGENT project was presented at the plenary session. Contacts have been established with other EGEE related projects. DILIGENT contributed along with other projects in providing feedback to the networking activities.		
December 2004			
#11	**Subscription to EGEE User Information Group**		
	DILIGENT partner is involved in the activity of the EGEE UIG.		
#12	**Joint technical meeting n.4**		Geneva, Switzerland
	Design of the DILIGENT Content and Metadata Management Service with JRA1 team.		
#13	**Joint technical meeting n.4**		Geneva, Switzerland
	Organisation of DILIGENT testing activity with JRA1 testing team.		
#14	**Access to the EGEE prototype testbed**		
	Access to the EGEE prototype testbed has been granted to CERN to support the gLite study activity previously conducted mainly gathering information from various sources (available documentation, meetings, web site). Study of gLite components, development, testing, integration and plans has been performed by CERN and presented to DILIGENT partners at the end of January 2005.		
January 2005			
#15	**NA4 Wiki Site contributions**		
	Registration to the EGEE NA4 Wiki site and upload of information related to DILIGENT.		
February 2005			
#16	**Subscription to gLite mailing lists**		
	DILIGENT partners subscribed to various gLite technical mailing lists.		
#17	**Access to EGEE prototype testbed**		
	Access to the EGEE prototype testbed has been granted to a restricted number of DILIGENT members for experimentation purposes. Experimentation has started.		
#18	**1st EGEE review**		Geneva, Switzerland

	DILIGENT scientific coordinator and CERN partners attended the first EGEE review.	
March 2005		
#19	**EGEE JRA1 contacts**	Geneva, Switzerland
	EGEE JRA1 coordinators have been informed about DILIGENT plans for deployment and experimentation.	
April 2005		
#20	**Joint training event**	Athens, Greece
	"Grid Technologies for Digital Libraries" training has been organised as a joint effort between DILIGENT, DELOS and EGEE.	
#21	**3ʳᵈ EGEE Conference**	Athens, Greece
	DILIGENT partners attended the EGEE conference. DILIGENT project was presented at the plenary session. Dissemination stand was set up by the DILIGENT dissemination WP team.	
#22	**Joint technical meeting n.5**	Athens, Greece
	Discussion with the coordinator of the EGEE American Federation on deployment issues.	
#23	**Joint technical meeting n.6**	Athens, Greece
	Discussion with JRA1 team on various technical issues and deployment planning.	
#24	**EGEE coordinators contacts**	
	EGEE coordinators have been informed about DILIGENT plans for deployment and experimentation.	
#25	**Collaboration with EGEE test team**	
	Investigation on opportunities for collaboration between DILIGENT partners and EGEE test team.	
May 2005		
#26	**Contacts with EGEE NA4 coordinator**	
	Collaboration in defining a support strategy for EGEE related projects.	
Design and Experimentation (M9-M22)		
May 2005		
#27	**Bugs submission**	
	Bugs submission to bugs tracking tools and gLite discussion list	
#28	**Requirements submission**	
	Middleware requirements submission to EGEE requirements database	
June 2005		
#29	**gLite deployment tutorial**	Geneva, Switzerland
	Participation of DILIGENT members in gLite deployment tutorial	

#30	**Collaboration with EGEE testing team**	Geneva, Switzerland
	One DILIGENT member worked for one week with EGEE testing team	
#31	**TestManager**	
	Adoption of EGEE testing team templates for DILIGENT gLite tests	
#32	**Bugs submission**	
	Bugs submission and discussion on gLite discussion list glite-dicuss@cern.ch	
July 2005		
#33	**Testing, integration and deployment**	Geneva, Switzerland
	Meeting with JRA1 Integration Leader Alberto Di Meglio about EGEE testing, integration and deployment model and its possible usage by DILIGENT	
September 2005		
#34	**EGEE PPS infrastructure**	Geneva, Switzerland
	Meeting with EGEE SA1 PPS responsible Nicholas Thackeray about DILIGENT access to EGEE PPS infrastructure	
#35	**EGEE PPS infrastructure**	
	Access to EGEE PPS infrastructure was granted to all DILIGENT partners	
#36	**TestManager**	
	Java testing tool TestManager, used by JRA1 testing team, was tested. Feedback was given.	
October 2005		
#37	**4th EGEE Conference**	Pisa, Italy
	General plenary presentation. Technical presentation of DILIGENT services and gLite usage.	
#38	**4th EGEE Conference**	Pisa, Italy
	DILIGENT-EGEE session on DILIGENT exploitation of gLite functionality.	
November 2005		
#39	**Meeting with EGEE Technical Director**	Geneva, Switzerland
	Meeting with EGEE Technical Director Erwin Laure to discuss the goal of the Technical Coordination Group (TCG) and DILIGENT participation in this group	
#40	**EGEE NA3 internal review**	
	Coordination of DILIGENT participation in the EGEE NA3 internal review.	
#41	**Meeting with EGEE NA4**	Geneva, Switzerland
	Meeting with NA leaders - Vincent Breton, Massimo Lamanna, Roberto Barbera and Cal Loomis - to discuss the future collaboration between DILIGENT and EGEE NA4	
#42	**EGEE Middleware Security group meeting**	Amsterdam, Netherlands
	DILIGENT joined and participated in the meetings organised by the Middleware	

	Security working Group and VOMS developers.	
#43	**EGEE Joint Policy Security Group meeting**	Amsterdam, Netherlands
	DILIGENT joined and participated in the meetings organised by this working group promoted by EGEE and LCG.	
December 2005		
#44	**2nd EGEE EU review**	Geneva, Switzerland
	DILIGENT partners attended the EGEE review	
#45	**Meeting with EGEE NA4**	
	Meeting with NA4 people (Charles Loomis and Birger Koblitz) to discuss the DILIGENT requirements and the DILIGENT involvement in TCG workgroups: MDM and SDJ.	
January 2006		
#46	**EGEE NA4 Generic Applications meeting**	Catania, Italy
	DILIGENT was presented to the Generic Applications community.	
#47	**EGEE All-activity meeting**	Geneva, Switzerland
	DILIGENT partners attended the EGEE all activity meeting.	
#48	**Meeting with EGEE NA4**	
	DILIGENT joined MDM and SDJ workgroups. Discussion with Cal Loomis about DILIGENT participation in JP workgroup.	
February 2006		
#49	**EGEE TCG JP workgroup meeting**	Geneva, Switzerland
	DILIGENT partners joined JP workgroup. First face-to-face meeting. DILIGENT requirements were presented.	
#50	**Integration and ETICS**	Geneva, Switzerland
	Meeting with ETICS Coordinators to discuss about DILIGENT integration activity and ETICS usage	
March 2006		
#51	**Participation to the EGEE User Forum**	Geneva, Switzerland
	DILIGENT partners attended the EGEE User Forum and presented the DILIGENT project.	
#52	**EGEE TCG SDJ workgroup meeting**	Geneva, Switzerland
	DILIGENT partners attended the TCG SDJ meeting and presented DILIGENT requirements.	
#53	**EGEE Middleware Security Group meeting**	Geneva, Switzerland
	Update on current global security architecture work	
April 2006		
#54	**Meeting with EGEE NA4**	Geneva, Switzerland
	Meeting with NA4 leader Cal Loomis to discuss: the work done by TCG, the work	

	done by DILIGENT in the TCG workgroups, DILIGENT requirements for gLite3.0 and the new organisation of NA4, need for DILIGENT MoU.	
May 2006		
#55	**Meeting with EGEE technical coordination**	Geneva, Switzerland
	Meeting with EGEE Technical Director E.Laure and JRA1 Leader C.Grandi to present DILIGENT usage of gLite and discussion on major requirements.	
#56	**EGEE TCG JP workgroup meeting**	Geneva, Switzerland
	DILIGENT partners participated in the TCG JP workgroup meeting.	
#57	**Meeting with EGEE SA3 leader**	Geneva, Switzerland
	Meeting with EGEE SA3 leader M.Schulz to discuss the requirements for connecting the DILIGENT infrastructures with the EGEE production (and pre-production) infrastructures.	
#58	**Meeting with EGEE SA3**	Geneva, Switzerland
	Meeting with L.Field to discuss the interoperability issues of connecting the DILIGENT infrastructures with the EGEE production and pre-production infrastructures.	

Table 1: DILIGENT-EGEE interactions

Motivations for the above interactions are listed in the following table:

	Interaction	**Motivations**	**Areas**
Hands-on Technology (M0-M8)			
#1	1st EGEE Conference	Disseminate/Demonstrate	Networking
#2	DILIGENT Kick-off Meeting	Learn Disseminate/Demonstrate	Networking Middleware
#3	Joint technical meeting n.1	Learn	Middleware
#4	EGEE All-activity meeting	Learn	Middleware
#5	Joint technical meeting n.2	Learn	Middleware
#6	Subscription to EGEE user requirements database	Provide Feedback	Middleware Networking
#7	ARDA workshop	Learn	Middleware
#8	Joint technical meeting n.3	Learn Implement	Middleware
#9	Joint technical meeting n.3	Learn Implement	Middleware
#10	2nd EGEE Conference and Concertation meeting	Disseminate/Demonstrate Learn	Networking Middleware

#11	Subscription to EGEE User Information Group	Provide Feedback	Networking
#12	Joint technical meeting n.4	Learn	Middleware
#13	Joint technical meeting n.4	Learn	Middleware
#14	Access to the EGEE prototype testbed	Learn Align	Middleware
#15	NA4 Wiki Site contributions	Disseminate/Demonstrate	Networking
#16	Subscription to gLite mailing lists	Learn	Middleware
#17	Access to prototype testbed	Experiment	Middleware
#18	1st EGEE review	Learn	Networking
#19	EGEE JRA1 contacts	Align	Middleware
#20	Joint training event	Learn Disseminate/Demonstrate	Networking Middleware
#21	3rd EGEE Conference	Disseminate/Demonstrate Learn	Networking Middleware
#22	Joint technical meeting n.5	Learn Deploy	Operations
#23	Joint technical meeting n.6	Learn Deploy	Middleware Operations
#24	EGEE coordinators contacts	Align	Networking
#25	Collaboration with EGEE test team	Learn Implement Deploy	Middleware Operations
#26	Contacts with EGEE NA4 coordinator	Provide Feedback	Networking
Design and Experimentation (M9-M22)			
#27	Bugs submission	Provide Feedback Implement	Middleware
#28	Requirements submission	Provide Feedback	Middleware
#29	gLite deployment tutorial	Learn Deploy Experiment	Operations
#30	Collaboration with EGEE testing team	Learn Deploy Experiment Implement	Middleware
#31	TestManager	Implement	Middleware
#32	Bugs submission	Provide Feedback Implement	Middleware

#33	Testing, integration and deployment	Learn Deploy	Operations Middleware
#34	EGEE PPS infrastructure	Align Experiment	Operations
#35	EGEE PPS infrastructure	Experiment Implement	Operations
#36	TestManager	Provide Feedback	Middleware
#37	4th EGEE Conference	Disseminate/Demonstrate Provide Feedback	Networking
#39	Meeting with EGEE Technical Director	Provide Feedback Align	Middleware
#40	EGEE NA3 internal review	Provide Feedback	Networking
#41	Meeting with EGEE NA4	Align Provide Feedback	Networking Middleware
#42	7th EGEE Middleware Security group meeting	Align Provide Feedback Implement Learn	Middleware
#43	EGEE Joint Policy Security Group meeting	Align Provide Feedback Implement Learn	Middleware
#44	2nd EGEE EU review	Align Learn	Networking
#45	Meeting with EGEE NA4	Align Provide Feedback	Middleware Networking
#46	NA4 Generic Applications meeting	Disseminate/Demonstrate	Networking
#47	EGEE All-activity meeting	Learn	Middleware
#48	Meeting with EGEE NA4	Provide Feedback Align	Middleware Networking
#49	EGEE TCG JP workgroup meeting	Provide Feedback Implement Learn Align	Middleware
#50	Integration and ETICS	Align Implement Learn	Operations Middleware

#51	Participation in the EGEE User Forum	Disseminate/Demonstrate	Networking
#52	EGEE TCG SDJ workgroup meeting	Provide Feedback Implement Learn Align	Middleware
#53	8th Middleware Security Group meeting	Implement Learn Provide Feedback	Middleware
#54	Meeting with EGEE NA4	Align Provide Feedback	Middleware Networking
#55	Meeting with EGEE technical coordination	Provide Feedback Align	Middleware
#56	EGEE TCG JP workgroup meeting	Learn Provide Feedback	Middleware
#57	Meeting with EGEE SA3 leader	Deploy Implement Learn	Operations
#58	Meeting with EGEE SA3	Deploy	Operations

Table 2: Motivations for DILIGENT-EGEE interactions

2　HANDS-ON TECHNOLOGY (M1 – M8)

This chapter describes in detail the interactions made between project month 1 and project month 8: from September 2004 to April 2005. Many of them focussed on the need of gaining a deep understanding of the gLite middleware DILIGENT has planned to build on.

2.1　Interactions related to Middleware

In this first phase of the DILIGENT project significant efforts were dedicated to the study the gLite middleware, the EGEE documentation and to understand how to better exploit it for satisfying the project's objectives. Several actions were carried out to acquire, as early as possible, familiarity with this middleware since a deep knowledge of it was deemed essential for the start-up of the DILIGENT design and development activities.

A first major input to this learning process came from the publication of the "EGEE Middleware Architecture and planning – Release 1" on 18th August 2004 (one month before the official starting of the DILIGENT project activities). This document describes the architecture of the EGEE middleware, named gLite, by means of a service oriented architecture. As specified by the authors, this has to be considered "a living document and we expect modifications of the proposed architecture as we gain experience with practical implementations, feedback from our users and evolution of the requirements". The first DILIGENT – EGEE interaction meeting was dedicated to discuss and clarify our understanding of this document.

The publication of the "Design of the EGEE Middleware Grid Services" report on 30th September 2004, represents another milestone in the comprehension of the EGEE middleware plans. This document describes the design of the external interfaces of gLite. A detailed specification of the internal functionality of the services, their dependencies, and deployment models, that are key aspects for the implementers of both the gLite services and the application services based on the gLite middleware, are beyond the scope of the cited document. These have been presented in subsequent documents. Many of these documents have been studied and discussed with EGEE representatives in a series of meetings before and after their official publications. The close interaction with the JRA1 members has thus enabled the DILIGENT team to deal with changes and evolutions of the EGEE Middleware. Moreover, it allowed completing successfully the DILIGENT functional specification before the first release of the gLite middleware that has been made available in April 2005 with a seamless documentation, installation instructions, and release notes.

During this period another important learning activity has been the experimentation of LCG middleware first and gLite later. This has highlighted many interesting operational aspects and it opened several questions discussed during the joint meeting.

All the interactions mentioned above are described in details in the next sections.

2.1.1　Joint Technical Meetings

Several joint technical meetings were held in this first 8 months of the DILIGENT project activity. Most of them have been decided along the way as follow-up of informal discussions or as a consequence of specific necessities raised during the DILIGENT design phase.

In order to reduce the travel expenses most of these meetings were held jointly with other events (e.g. EGEE conferences). Some of them were organised as "brainstorming" meetings dedicated to specific topics. Such meetings revealed to be very stimulating frameworks for merging and comparing different experiences and have been particularly appreciated by the participants who have asked to repeat these experiences also in the future.

Below is presented the list of the meetings held, and for each of them a brief description is reported. Other similar meetings are being planned.

DILIGENT Kick-off Meeting (#2) and Joint technical meeting n.1 (#3), Pisa, 8th September 2004

Topic/Purpose: Present DILIGENT objectives, EGEE organisation and gLite status. Discuss about specific requirements of the DILIGENT application.

Participants: Representatives of DILIGENT WP1.2 and WP3.1, Frederic Hemmer (EGEE middleware activity leader), Fabrizio Gagliardi (EGEE project leader).

Total number of participants: 10

Result: This first face-to-face meeting between the representatives of the two projects offered the opportunity to clarify the objectives of the DILIGENT project and the EGEE middleware plans and timing. It also provided the occasion for starting a discussion on specific DILIGENT technical requirements. In particular, this meeting, by providing a framework for presenting the DILIGENT requirements related to the Grid middleware, contributed to the identification of the following critical issues:

- Monitor VOs resources;
- Create VDL authorization environment;
- Deploy and Create VDL services instances;
- Prepare Grid sites with a set of selected software packages in order to host a service;
- Submit DAG of jobs;
- Monitor jobs status;
- Retrieve the job output;
- Release a set of resources;

Moreover the following distinguishing characteristics were identified:

- Data security (access policy, encryption, watermarking, etc.);
- Data Management (not only file management, but also structured directories support in the catalogues);
- Resources and services discovery on demand.

Documentation: Meeting minutes are available on the project website in the area restricted to the project participants[1].

Joint technical meeting n.2 (#5), Pisa, 30th September – 1st October 2004

Topic/Purpose: Brainstorming on the design of the DILIGENT Collective Layer Services and its relationships with the underlying Grid infrastructure.

---

[1] DILIGENT working area restricted to projects' members: http://dlib.sns.it/bscw/bscw.cgi/

Participants: CNR staff members, Frederic Hemmer and Miron Livny (coordinator of the American federation in EGEE and member of the middleware team).

Total number of participants: 7.

Organisation: In the afternoon of the first day CNR presented shortly the history of DLs in order to describe their recent evolution and their long term vision. A state-of-the-art system, OpenDLib, partially developed in the SCHOLNET project, was also presented as an example of the current technology. During the second day the CNR participants illustrated a use story from both the user point of view and the DILIGENT point of view. This story was mainly focussed on:

- the creation and management of a VDL
- the use and enrichment of a VDL

The story was used as a basis for presenting the CNR point of view and concerns regarding the DILIGENT expectations on:

- the OGSA-compliant middleware
- the EGEE (gLite) infrastructure to support jobs execution and data management in a reliable environment

and to introduce a detailed analysis on the issues related to the dynamic service deployment and to the security requirements.

The story stimulated a discussion and the EGEE representatives responded to the issues raised by making comments, remarks and by providing clarifications and suggestions.

	Presentation	*Speaker*
30 Sept	DL requirements	Donatella Castelli(CNR)
	An example of DL system: OpenDLib	Pasquale Pagano(CNR)
1 Oct	Use story	Pasquale Pagano(CNR)
	Brainstorming	All

Table 3: Agenda of the collective layer meeting

Results: The EGEE representatives got a better understanding of some of the specific DL requirements and clarified to the CNR team various aspects related to security, deployment, communication, rules, etc. of the EGEE middleware.

The participants to the meeting decided to repeat this experience also in the future since they considered this type of meetings very fruitful. A plan was established to organise a series of joint meetings, each focussed on other specific DILIGENT aspects and functionalities.

Documentation: The DILIGENT objectives, the proposed technical solution and the use story are described in a document "DILIGENT-EGEE interaction document n.1" that was distributed to all the participants before the meeting.

A new document "DILIGENT-EGEE interaction document n.2", which describes how DILIGENT will exploit the Grid middleware, was prepared as an outcome of this meeting.

This document was distributed to all the other partners and it was used as an input for the Joint Technical Meeting #3.

The technical documentation, the presentations and the minutes of the meeting have been published on the project website and they are accessible to all DILIGENT members.

Joint technical meeting n.3 (#8, #9), Rome, 19th -20th November 2004

Topic/Purpose: Brainstorming on DL Creation and Management and Search Service.

Participants: Miron Livny, Donatella Castelli (DILIGENT scientific co-ordinator), Pasquale Pagano (DILIGENT technical co-ordinator), representatives of DILIGENT WP1.4 and WP3.1. Total number of participants: 14.

Organisation: The workshop had two main sessions, one dedicated to the discussion of issues related to the realization of the DL Creation and Management services and another one which focused on aspects related to the Search service. Both sections started with an introduction on the requirements on the two classes of services made by the corresponding WPs leaders. This introduction raised a number of issues that were discussed in the brainstorming sessions.

	Presentation	*Speaker*
19 Nov	Introduction to DL Creation and Management	Pasquale Pagano (CNR)
	Brainstorming session 1	CNR, UoA, CERN
20 Nov	Introduction to the Search Service requirements	George Kakaletris, Yannis Ioannidis (UoA)
	Brainstorming session 2	All

Table 4: Agenda of the DL creation and management and search service meeting

Results: At the time almost no documentation about the gLite middleware was publicly available, therefore the initial DILIGENT vision was prepared relying on previous results produced by Grid infrastructure projects (e.g. DataGrid) and on experiences made by other Grid application (e.g. Grace). The meeting enabled the DILIGENT team to improve its understanding of the EGEE middleware. In particular, the information acquired about the status and the capabilities of gLite and the Condor framework allowed the DILIGENT team to revise and refine its vision on how the DILIGENT infrastructure will exploit the underling Grid functionalities provided by the gLite middleware.

Documentation: An initial concrete technical vision was circulated to all participants as input for the meeting discussion. The objective of this document was to state a common background vision of the project in order to achieve an agreement among the partners involved in the development of the DILIGENT system.

The document, which emphasizes at a high level some technical aspects related to the whole DILIGENT project, results from:

- the analysis on the status of the art of the Grid technologies and infrastructure;
- the outcome of the DILIGENT-EGEE joint technical meetings n.1 and n.2.

By presenting the relationship between DILIGENT and the Grid infrastructure the document first introduces the context where DILIGENT operates, then it discusses some issues related to the design and the implementation of the software.

Also for this meeting, the technical documentation, the presentations and the minutes were published on the project website and they are accessible to all DILIGENT members.

Joint technical meeting n.4 (#12), Geneva, 16th December 2004

Topic/Purpose: Metadata and content management are central issues in the digital library domain. Metadata are especially of central importance for efficient information access. Advanced and manifold formats have been developed and standardized for this purpose. Special requirements are imposed to metadata management by the digital library infrastructure planned in DILIGENT like covering a wide variety of content objects.

On the other side, in the development of the gLite infrastructure there is also a notion of data and metadata management that is considerably different from the understanding and practices in the digital library community.

The technical meeting organised at CERN aimed at discussing data management related functionalities provided by EGEE, as well as the DILIGENT consortium providing feedback about desired functionalities.

For the DILIGENT members it was the purpose of this meeting to learn how the EGEE infrastructure can support the management of metadata and what their approaches are for this task. For the EGEE representatives the purpose of the meeting was to better understand the requirements imposed by having digital libraries and especially DILIGENT as an application of the EGEE infrastructure.

Participants: Pasquale Pagano, representatives of DILIGENT WP1.3 and WP3.1, Frederic Hemmer, Erwin Laure (EGEE middleware activity leader deputy) and Peter Kunszt (leader of EGEE Data Management activity).

Total number of participants: 10.

Organisation: For the meeting several presentations were organised (see the agenda in Table 5). The meeting itself was spread over one complete day (16th December, 2004) and was attended by representatives from CERN, Fraunhofer IPSI, CNR, ETH Zürich and the University of Strathclyde.

	Presentation	*Speaker*
	Introduction to DILIGENT project	Pasquale Pagano (CNR)
	Introduction to EGEE/JRA1	Frederic Hemmer(CERN)
AM	Coffee break	
	Presentation on EGEE data/metadata management	Peter Kunszt (CERN)
	Presentation on Metadata Management	Bhaskar Mehta (FhG)
	Presentation on Content Management	Sören Balko(ETH)
Lunch		
PM	Data/metadata	All (DILIGENT/EGEE)

Discussions on Distributed Data Management, Services for Metadata, EGEE in other Metadata Contexts	All (DILIGENT)

Table 5: Agenda of the metadata meeting

The meeting started with a common session in the morning introductory to DILIGENT and EGEE. This was followed by a more focussed session for discussion on data management. EGEE representatives provided an overview of the overall SRM (storage and resource management) sub system of EGEE and presented a time line for the planned functionalities. There after, DILIGENT representatives presented their view on data management and required functionalities for DILIGENT. This was in the form of two presentations from ETH and Fraunhofer IPSI about content management and metadata management respectively. The afternoon session was organised to encourage more detailed discussion on specific issues and clarifications sought by the DILIGENT consortium.

Results: Currently, the planned metadata support in EGEE is at the level of metadata support provided by operating systems. Metadata are associated to files and directories (cf. objects and collections in DILIGENT). Metadata are generally represented in the form of key/value pairs, the values are atomic, i.e., they are not further structured. Directories may have different metadata schemes, one metadata scheme applies to the entire content of a directory. Metadata are either stored inline with files, or in separate metadata catalogues that refer to files and directories. Inline metadata cannot be used for efficient search but may be updated more efficiently, while metadata catalogues can be searched. This basic infrastructure may be used by DILIGENT where appropriate, i.e. when objects are represented as files and collections as directories and flat key/value pairs suffice. For finer grained metadata at the level of objects and more structured metadata, DILIGENT's metadata repository will be required.

Some of the issues discussed in this meeting included:

- Distribution of Metadata (Querying, Updating);
- Association of Metadata with distributed data (Consistency, ...) e.g. at what granularity it is possible to talk about data/objects (files with URL, or finer grained);
- Using databases–like approaches for data management in DILIGENT;
- How does EGEE currently deal with typical metadata (e.g. access-rights, ...);
- Services for Metadata (Production, Analysis, Maintenance);
- Versioning of Content (CVS-like solutions).

The results of the meeting can be summarized as follows:

- Presentation of the current status of EGEE and its SRM related features;
- Presentation of requirements from DILIGENT;
- Clarifications regarding technical issues raised by DILIGENT partners listed above.

Documentation: All the presentations are on the project website in the area restricted to the project members and publicly available at CERN online agenda[2].

Joint technical meeting n.4 (#13), Geneva, 16th December 2004

---

[2] http://agenda.cern.ch/fullAgenda.php?ida=a045520

Topic/Purpose: This meeting has been set up with the main goal of passing knowledge from EGEE to DILIGENT about the process implemented by the EGEE project to ensure the quality of the software released. In particular integration and testing techniques and tools used, testing and integration teams organisation, lesson learned from past experience.

Participants: Representatives of WP1.7, Maite Barroso Lopez (EGEE testing/integration task manager), Leanne Guy (EGEE testing team leader) Andrea Di Meglio (EGEE integration team leader).

Total number of participants: 7.

Organisation: The meeting was held on 16th December 2004 and was attended by representatives from CERN, 4DSOFT and ENG.

	Presentation	*Speaker*
AM	Introduction to DILIGENT project	Pasquale Pagano (CNR)
	Introduction to EGEE/JRA1	Frederic Hemmer (CERN)
	Coffee break	
	Testing brainstorming session	All (CERN/DILIGENT)
Lunch		
PM	DILIGENT testing	Istvàn Forgacs (4DSoft)
	EGEE testing	Leanne Guy (CERN)
	EGEE integration	Alberto Di Meglio (CERN)

Table 6: Agenda of the testing meeting

The meeting started early in the morning. A common session with the Data Management teams was meant to introduce DILIGENT and EGEE middleware activities to all participants. After the coffee break the test teams (from DILIGENT and EGEE) met to discuss about the main issues of testing in a Grid environment.

In the afternoon a series of presentations took place. Leanne Guy and Alberto Di Meglio presented the current implementation of the integration and testing techniques in EGEE. Istvàn Forgacs described the competencies of 4Dsoft, the testing phases of a typical large project and the core ideas of the EBIT approach to testing.

Results: 4DSOFT presented their initial ideas for DILIGENT testing. EGEE representatives presented their current testing/integration strategy, the lessons learned from their past experience and guidelines about how to orient testing/integration activities in distributed environments. First plans for DILIGENT testing were made. Some issues raised at the meeting and related outcome are detailed here:

- How to manage testing in a distributed and multi-paradigm programming environment: EGEE developers adopt several programming languages, which makes testing more difficult. DILIGENT should adopt a unique language, i.e. Java, unless specific performance problems require other programming languages, e.g. C or C++;
- What kind of testing tools are used by EGEE and can be adopted by DILIGENT:
  - o QMTest: Test suite used to run all their tests;

- o Jalopy: For coding analysis;
- o JUnit, CppUnit: Unit tests;
- o Clover, gCov: Coverage
- What kind of testing can be made:
  - o use 4D Soft own made testing tools;
  - o implement the category-partition method in all phases of the testing process;
  - o propose the EBIT approach to testing;
  - o unit testing: black box and white box testing by each programmer;
  - o integration and component testing: black box testing by each programmer;
  - o system testing (installation and configuration testing): by the testing team;
  - o acceptance testing: by the testing team;
  - o regression testing: by the testing team;
- How to implement the infrastructure (either for testing and development):
  - o pay attention to infrastructure configuration problems like external access, firewall configuration, etc;
  - o provide at each site a system administrator (can be just part-time) responsible for system maintenance and frequent installation of middleware and other applications;
  - o three infrastructures are needed (development, testing and production) and should include at least three sites with three nodes each (currently a node is a pool of at least three machines because some services are incompatible in the same PC);
- How to interact with EGEE for bugs reporting related to gLite middleware. A new bug reporting tool should be used instead of the EGEE one;
- What kind of internal organisation to propose at DILIGENT Technical Committee:
  - o Organisation of an initial tutorial for developers;
  - o A web based tool and guidelines:
    - ▪ to keep track of the bugs discovered during testing;
    - ▪ to report the status of testing activities toward the Steering Committee and the Commission;
    - ▪ to interact with developers (in DILIGENT and in EGEE) for bugs discovered by other testing activities (system testing, acceptance testing).

Documentation: Presentations are published on the project website in the area restricted to the project participants and publicly available at CERN online agenda[3].

Joint technical meeting n.5 (#22), Athens, 19th April 2005

Topic/Purpose: Taking into account that DILIGENT has scheduled the deployment of its own development Grid infrastructure in May 2005, and that a concrete experimentation with the ESA-ESRIN partner are being designed, the infrastructure deployment and tuning and the application design were the main discussion topics of this meeting.

Participants: Representatives of DILIGENT WP1.2 and Miron Livny.

Total number of participants: 6.

Results: The main issues discussed during this meeting were:

---

[3] Link to the meeting agenda and presentations: http://agenda.cern.ch/fullAgenda.php?ida=a045520

- Deployment and tuning of the infrastructure (e.g. how many instances of a service are required to satisfy the DILIGENT needs); how to allocate services on nodes (i.e. grouping of different compatible services on the same node vs. hosting each service on a different node); etc.
- DAG of jobs vs. storm of jobs.

Taking into account that no experimental data is already available and that the correct tuning of some of the typical applications of the DILIGENT scenario requires explicit experimentations and analysis of the results, this meeting allowed to identify with the expertise provided by Miron Livny the following first initial set of experiments: a) DILIGENT will initially deploy a basic infrastructure to support mainly job submission, execution, and monitoring; b) DILIGENT will experiment the same scenario with the two approaches, DAG and storm, (with an incremental number of jobs) and will collect results. The analysis of these results will provide feedback and will be the foundation for future technical meetings.

Documentation: No initial documentation has been produced for this meeting. The results of the experiments will be posted on the project website.

Joint technical meeting n.6 (#23), Athens, 21st April 2005

Topic/Purpose: Brainstorming on gLite information service, batch system, and DILIGENT applications.

Participants: Representatives of DILIGENT WP1.2, Frederic Hemmer, Erwin Laure, Peter Kunstz, Steve Fisher (EGEE Information Services activity leader).
Total number of participants: 8.

Results: The discussion covered aspects related to:
- the integration of the information system (R-GMA) provided by the current and future releases of the gLite middleware in the DILIGENT information system;
- the batch systems (LSF, PBS, Condor) with the tentative to emphasize their main characteristics and limitations with respect to the DILIGENT needs;
- the job definition language (JDL) in order to understand what would be the best way to represent DILIGENT needs.

As a result of this meeting it was decided that a set of experiments will be deployed on the DILIGENT Grid infrastructure and the results will be discussed with the EGEE JRA1 team. A new mailing list will be set-up and maintained by the EGEE JRA1 team to improve and enhance the EGEE interactions with gLite user communities and DILIGENT application developers in particular.

Documentation: No documentation has been produced as an input for this meeting. The results of the experiments will be posted on the project website.

2.1.1.1 Outcome

The meetings among a restricted number of people very knowledgeable in their respective fields have proved to be extremely useful for cross-fertilization among the DLs and Grid domains. Important improvements in better focussing the problems and in designing and refining solutions have been achieved as a result of these meetings. As a consequence some of the initial design choices made by the DILIGENT partners have been revised while

others have been consolidated. These meetings have also provided a framework where researchers and developers from the DLs and Grid domains had the opportunity to discuss and compare their experiences. This exchange has fostered further collaborations that extend beyond the boundaries of the EGEE and DILIGENT projects. As a side effect of these joint meetings, a number of important aspects that DILIGENT only partially covers, due to the limited effort allocated to them, were highlighted. In order to face some of them, two proposals, BELIEF (Bridging Electronic Infrastructure to Expanding Frontiers) and ECTICS (eInfrastructure for Testing, Integration and Configuration of Software), have been prepared by some of the DILIGENT partners and submitted under the FP6, Research Infrastructure Action 3.2.3 *"Communication Network Development – eInfrastructure – Consolidating Initiatives" call of the* FP6.

The EGEE JRA1 team has scheduled one release of gLite per month, in order to provide bug fixing and mandatory unpublished functionalities, and one major release every three months, in order to provide steps forward to the current release with respect to the services integration and to the provided functionalities. Given this very dynamic situation of the gLite middleware, the close technical cooperation and the interactions between the DILIGENT and EGEE projects represent key and necessary means for the detailed design and realisation of the DILIGENT services, and their successful exploitation of the gLite middleware.

2.1.2 EGEE/gLite Study

Since the very beginning of the project, DILIGENT partners have started collecting information on the status of the EGEE middleware, i.e. gLite. A more detailed study on gLite available components and development process has been performed by CERN partner. This activity was performed between October 2004 and the end of January 2005. Results were presented to the DILIGENT partners during the technical meeting held in Hall in the first days of February. The goals of this study were to gain a basic understanding of gLite architecture, design, functionalities, installation process and testing strategy; make gLite documentation available to the partners; understand how DILIGENT adoption of the middleware will happen.

At that time, gLite middleware was still under development, the documentation available (user guides, APIs, installation guides) was obviously incomplete and interactions with EGEE representatives from the middleware, testing and integration teams were needed. In this way a better understanding of what was expected to be available on gLite release 1 has been achieved, as well as knowledge on the functioning and organisation of EGEE development, integration and testing activities.

As an on-going procedure, during this period, results and milestones on EGEE activity were communicated to all DILIGENT partners and technical issues were raised and discussed between technical partners. Several methods were used to pass this information:

- Emails: for announcements, communications, etc;
- DILIGENT repository: for upload of report, minutes and interim work documents;
  - o http://dlib.sns.it/bscw/bscw.cgi/0/7777
  - o http://dlib.sns.it/bscw/bscw.cgi/0/14733
  - o http://dlib.sns.it/bscw/bscw.cgi/0/11721
- Internal website: for publishing the gLite tests and the results achieved;
  - o Current link: https://uimon.cern.ch/twiki//bin/view/DILIGENT/WebHome
  - o Future link: http://diligentproject.org/content/section/7/129/
- Conference calls: to directly discuss on certain topics;

- Presentations: to give an overall vision of the work done and its results.
  - o http://dlib.sns.it/bscw/bscw.cgi/d11931/gLite_status_experimentations.ppt

The study of the gLite middleware was performed in several stages. Looking to Figure 2 it is possible to understand which gLite components were evaluated, which versions were used and when those tests took place. For a better understanding of the general architecture of gLite and its components more information is provided in Appendix A.
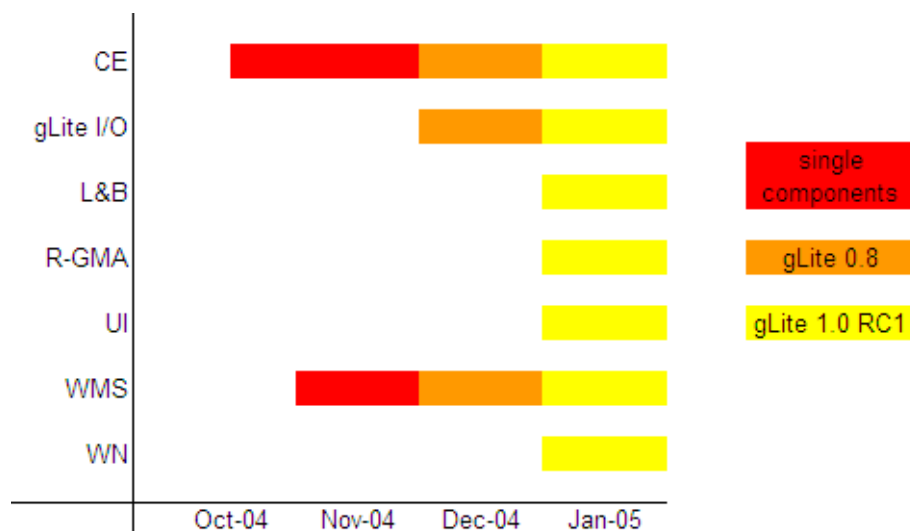

Figure 2: gLite components tested

Table 7 lists in more detail the kinds of tests (installation/usage) done on each component during this period. The three colours (red, orange and yellow) also identify three different stages of the evaluation activity (referring to Figure 1).

On the first stage, between October and November 2004, only two components were evaluated as independent packages and not as parts of a unique gLite release. Different sources of information were used to collect documentation and code: EGEE-JRA1, IFNF and LCG. These first tests were basically installations tests, to understand the structure of each component.

	Single components		gLite 0.8		gLite 1.0 RC1	
	installation	usage	installation	usage	installation	usage
CE	√		√	√	√	√
gLite I/O			√		√	√
L&B					√	
R-GMA					√	
UI					√	
WMS	√		√		√	√
WN					√	

Table 7: Type of tests by component

On the second stage, during the month of December, gLite study continued focusing on the usage of gLite version v0.8. The components that were evaluated were the CE and the WMS plus gLite I/O. The main difference from the October/November tests was the deployment method used since in December the official gLite installation procedure was made available. The same components previously evaluated were chosen in order to make use of the previously acquired experience to better understand the new deployment mechanism. Besides installation, simple usage of the Computing Element component was also tested, trying to execute the functionalities available at that time.

On the final stage, in January 2005, when EGEE was getting closer to release gLite 1.0, broader and more intensive experimentations were done on gLite version 1.0 RC1. Almost all the components available at that time were installed: WMS, CE, UI, gLite I/O, WN, R-GMA, L&B. Simple tests were implemented for some of them: WMS, CE, UI, gLite I/O. These tests were mainly based in the usage of command line interfaces provided. Two examples are given here:

- Computing Element

```
[prodrigu@pcedg002]~% voms-proxy-init --confile /opt/glite/etc/vomses
[prodrigu@pcedg002]~% globus-job-run pcedg002 /bin/pwd /home/egtest001
```

- Workload Management System

```
[prodrigu@pcedg002]~% voms-proxy-init --confile /opt/glite/etc/vomses
[prodrigu@pcedg002]~% export X509_USER_PROXY=/tmp/x509up_u17395
[prodrigu@pcedg002]~% globus-url-copy file:///afs/cern.ch/user/p/prodrigu/dummy
                      gsiftp://pcedg002.cern.ch/tmp/testitup
[prodrigu@pcedg002]~% voms-proxy-init --confile /opt/glite/etc/vomses egtest
[prodrigu@pcedg002]~% glite-job-submit --config-vo testing-tb-voconfig.conf pma_test.jdl
[prodrigu@pcedg002]~% glite-job-status https://lxb1419.cern.ch:9000/UVcbdNtcDp7dkvfi7aTE0A
```

The above mentioned proof of gLite performance and capabilities with respect the DILIGENT requirements were conducted mainly by CERN partner, with the participation of CNR-ISTI. Figure 3 represents which partners were involved in this activity, where they have performed the tests and when.
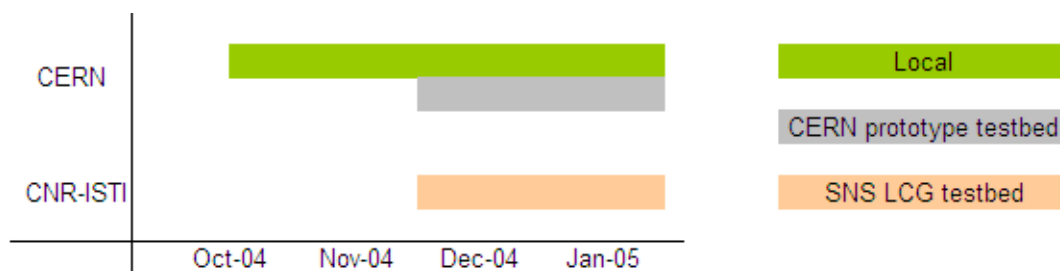

Figure 3: Tests execution environment

With respect to CERN activities, the majority of the gLite tests were done on single machines, not belonging to any testbed. At that beginning it was important to execute the tests on local machines since most of them were mainly installation tests and therefore a machine with administrator privileges was needed. Later, when the tests were then focused in running/using gLite rather then installing it, EGEE gLite prototype testbed was used. The usage of this testbed was agreed with EGEE middleware team and access was granted to a restricted number of technical partners (interactions #14, #17). The availability of the testbed allowed DILIGENT to easily execute more complex tests since there was no need for installing the required components.

At the same time, CNR used another testbed available at Scuola Normale Superiore (SNS), running LCG2 middleware. The tests performed focussed on the usage of the UI (installing X509 personal certificate and starting proxy certificates), job submission, and data management. The focus was on understanding Grid middleware capabilities.

2.1.2.1   Outcome

The major achievement of this activity was the fact that DILIGENT partners proved how to exploit/adopt the current gLite release capabilities/functionalities in DILIGENT. Knowledge on the functioning and organisation of EGEE development, integration and testing activities was also extremely useful since it provided the necessary information for the planning of DILIGENT deployment and development activity.

However one aspect has to be highlighted. gLite middleware was evolving extremely quickly during this period (new components, functionalities, bug fixes). This was a period of preparation for the release of the first official version of the middleware, which came to happen in March 2005. As expected before the official release of the software, the big number of changes, versions, bugs and the little documentation available made the study and testing work harder. Despite these difficulties, with the collaboration of EGEE technical teams, the study proved to be useful and the objective of obtaining a better understanding of what is gLite, what functionalities it provides and how it can be used/deployed was finally achieved.

2.1.3   gLite Experimentation

An activity dedicated to the experimentation of the gLite middleware was initiated in order to get as early as possible experience on its usage. The first objective of the experimentation was related to the testing of the features and the functionalities provided by gLite in the different areas of the middleware: job submission, data management, and information service. Experiments in the first two areas were done in order to gain confidence with the middleware on functionalities of interest to DILIGENT. The information service experiment was conducted in order to study the interaction between the DILIGENT information system (DIS) and the gLite information system (RGMA). Finally, also the security aspects were investigated.

The experimentation activity started as soon as gLite release 1 was made available. It is important to emphasize that the first release of the gLite middleware has been made available in April 2005. Before this date only limited experimentations could be done because the available infrastructures exposed only some of the features to be provided by gLite. Currently the experimentation is based on gLite official release and will follow the future releases as possible, taking into account their improvements and bug fixes.

The experimentation activity was performed by some of the DILIGENT partners (at present CERN, CNR-ISTI and ENG) in the EGEE prototype testbed. Here they had the possibility of conducting a set of initial experiments on the usage of gLite middleware.

In order to access this infrastructure they needed a personal certificate, issued by the local Certification Authority, then got a personal CERN account and, finally, installed a personal certificate into the gLite user interface (UI) and registered into the VO.

Job submission experimentation

gLite job management service is based on the Workload Management System (WMS). The WMS comprises a set of Grid middleware components responsible for the distribution and management of tasks across Grid resources in such a way that applications are conveniently, efficiently and effectively executed. The specific tasks that request

computation are usually referred to as "jobs". "Jobs" are defined using JDL files. The WMS is composed by three main components: WM, L&B and CE. The core component of the WMS is the Workload Manager (WM), whose purpose is to accept and satisfy requests for job management coming from its clients.

The first experiments focused on the following simple activities: job submission, job monitoring and job output retrieval. These tests represent the simplest and most typical sequence of job submission. An example of a JDL file for this kind of tests is as follows:

```
Executable = "/bin/echo";
Arguments = "Hello World";
StdOutput = "Message.txt";
StdError = "stderr.log";
OutputSandbox = {"Message.txt","stderr.log"};
```

The following tests focused in the management of input and output files. Regarding the usage of input files two tests were made. One using input files retrieved locally from the user interface and another where the input files were retrieved from a remote location, a predefined Storage Element. Concerning the output files, by default, they are retrieved and stored to a local place defined by the user. This was executed in all tests. The test for uploading and downloading the input and output files from and to a remote place was also developed. An example of a JDL file for this kind of test is as follows:

```
Type = "job";
JobType = "normal";
Executable = "t8.sh";
Arguments = "3 t8.in";
StdInput = "t8.in";
StdOutput = "t8.out";
StdError = "t8.err";
StorageIndex =  "http://lxb2028.cern.ch:8080/EGEE/glite-data-catalog-service-
fr/services/SEIndex" ;
InputData = { "lfn:/tmp/pma/t8.in" };
DataAccessProtocol= "gridftp";
OutputSandbox = {"t8.err","t8.out"};
InputSandbox = {"/afs/cern.ch/user/p/prodrigu/ui/jdls/t4.sh"};
```

Finally more complex tests were implemented and executed in order to exploit advance functionalities of the WMS and check if it suits specific DILIGENT requirements. These tests mainly focused in the execution of DAG jobs. DAG jobs allow the execution of a predefined sequence of jobs. All the possibilities for this type of job submission were analyzed. An example of a JDL file for this kind of tests is as follows:

```
 [
Type = "dag";
max_nodes_running = 5;
node_type = "edg-jdl";
nodes = [
nodeA = [
description = [
JobType = "normal";
Executable = "/bin/date";
StdOutput = "t5a.out";
StdError = "t5a.err";
OutputSandbox = {"t5a.out","t5a.err"};
];
];
nodeB = [
description = [
JobType = "normal";
Executable = "/bin/date";
StdOutput = "t5b.out";
StdError = "t5b.err";
OutputSandbox = {"t5b.out","t5b.err"};
];
];
```

```
dependencies = { { nodeA, nodeB }
};
];
];
```

Data Management experimentation

gLite data management services are based on four main components: storage management, file and replica catalogue, metadata catalogue and transfer scheduling. In this experimentation few tests were performed and only the file and replica catalogue (Single Catalogue) and one transfer scheduling service (File Transfer Service) were tested.

The Single Catalogue (SC) is responsible for the mapping between Logical File Names (LFN) and Site URLs (SURLs). The LFN is the key by which the users refer to their data. The SURL is a valid URL that specifies which Storage Element needs to be contacted to extract the data. SC tests were based in the usage of the provided command line interface (CLI). Through the CLI several simple interactions with the catalogue were tested:

- Create directories;
- Create files;
- Create symlinks;
- Remove files;
- Get file information.

The transfer scheduling services of gLite are responsible to securely transfer files between Grid sites. The transfer is performed always between two gLite Storage Elements having the same transfer protocol available to them (usually gsiftp). The File Transfer Service is responsible for the actual transfer of the file between the SEs. It takes the source and destination names as arguments and performs the transfer. The CLI provided was again used for executing the get and put functionalities

Information Service experimentation

gLite information and monitoring service is based on the R-GMA framework. R-GMA (Relational Grid Monitoring Architecture) is based on the GMA from the GGF (Grid Global Forum), which is a simple Consumer-Producer model.

This experimentation initially focused on the following topics: local installation of the various gLite R-GMA packages; access to R-GMA using the command line interface; tests of all the available commands; analysis of the standard tables (GLUE schema implemented); design and implementation of different types of producers and consumers.

Later, a more complex application, representing a simple prototype of the DILIGENT Information System (DIS), was designed and implemented. This application simulates the dynamic registration of new resource providers, the uploading of their state, and the monitoring by resource consumers. In particular, the application uses the most important functionalities provided by R-GMA made accessible through the Java API. The application can be described as decomposed in two sub-applications: the Producer and the Consumer. Moreover, the Producer application simulates all three types of resource providers supported by R-GMA. Over his execution the Producer randomly produces a set of resources; randomly selects a resource and change its state; and eliminates randomly existing resources. The generated resources update the Registry information using personalised tables.

Analogously, over his execution the Consumer monitors the available resources being notified about changes (addition of a resource, change into a resource state, and elimination of a resource).

Technical details about the experimentations already performed have been collected and made publicly accessible on the DILIGENT web site.

2.1.3.1   Outcome

Carrying out these first experiments was not simple. Many difficulties were encountered partially due to inexperience of the consortium in working with Grid technologies and partially to the novelty of the gLite middleware and the gLite documentation.

However, the progress done was quite important. It has been discovered and concluded that a closer integration of DILIGENT and gLite middleware is possible, profiting from a better exploitation of gLite functionalities.

The DILIGENT project is among the first external tester of the gLite middleware. This activity allowed starting the identification of the DILIGENT requirements not satisfied by gLite yet. These requirements, specific to the DILIGENT application, will be submitted to the attention of EGEE to be taken into consideration for gLite future developments.

2.2   Interactions related to Networking

EGEE networking activity is providing a number of instruments (web portal, training, requirements submission, collaboration area, etc.) to support EGEE user communities. The NA team is also constantly looking for interaction with users in order to better understand and satisfy their needs. DILIGENT is cooperating with the EGEE NA team in this direction.

At the same time DILIGENT needs to disseminate its plans, objectives and results to the EGEE and Grid communities. This awareness largely helps the establishment of contacts between the two projects as well as with other related initiatives and leads to:

- cooperation for a successful adoption and exploitation of the Grid technology by DILIGENT in the DLs domain;
- provision of feedback to EGEE activities regarding user support, middleware needs, infrastructure operation.

Interactions with EGEE are also needed in order to be able to maintain DILIGENT plans aligned with EGEE as both projects evolve.

2.2.1   Dissemination

One of the main objectives of DILIGENT is to promote cross-fertilization between the DLs and the Grid domains in order to foster synergies and advances in both areas. Networking with EGEE and other Grid related projects is essential in order to achieve this aim. Particular attention is paid by the members of the DILIGENT project to the participation to the EGEE events and to the Concertation Meetings organised by the EU since these provide an appropriate framework for disseminating the DILIGENT objectives, requirements and feedback to a large audience of key players in the addressed sectors.

The main reasons for these interactions are:

- dissemination of DILIGENT vision and plans to the EGEE and Grid community;
- information acquisition about EGEE status and plans;
- establishment of synergies with other initiatives.

DILIGENT members participated to the EGEE conferences and open meetings such us the all activities meetings and the EGEE review (interactions #1, #4, #10, #18, #21). These events provided the opportunities to increase reciprocal awareness about projects' objectives and directions, improving coordination and discerning opportunities for collaboration.

1st EGEE Conference (#1), Cork, 18th -22ndApril 2004

Objective: This first EGEE conference brought together all the partners from this project for the first time. There were plenary sessions and opportunities for the different EGEE Activities to meet in parallel sessions. Related projects, like DILIGENT, were invited to participate both to present their activity and to attend open project sessions.

Action: Donatella Castelli and Jessica Michael (respectively DILIGENT scientific and administrative co-ordinators) attended the conference as representatives of the DILIGENT project. D. Castelli gave a presentation "DILIGENT: A Digital Library Infrastructure on Grid Enabled Technology"[4] for disseminating the project objectives to the EGEE participants.

Result: Several links were established with EGEE partners, in particular with representatives of the NA4 activities, and with Miron Livny. Plans were made for future joint activities and meetings.

First Concertation Meeting on e-Infrastructure and 2nd EGEE Conference (#10), The Hague, 22nd-26th November 2004

Objective: The second EGEE conference was one of the events that took place in the Netherlands under the banner "European Leadership in e-Science and Grids" during the six months of the Dutch European presidency. Several EC Research Infrastructure Grid projects participated in this large event. During the Concertation meeting the partners of these projects had the opportunity to work together and to jointly investigate how each project can contribute to set-up an infrastructure serving the research projects.

Action: Donatella Castelli presented the DILIGENT objectives and plans[5] to the representatives of EGEE and other related projects.

A number of DILIGENT members attended the EGEE conference. Through the participation to the technical sessions they were informed about the last developments of the project.

Result: Synergies with projects related to EGEE were discussed. In particular, a plan for exploring joint activities with the SEE-Grid project was established. DILIGENT technical staff acquired a better understanding of the status of the EGEE infrastructure and gLite technology.

3rd EGEE Conference (#21), Athens, 18-22 April 2005

Objective: The 3rd EGEE conference was focussed on discussing the significant achievements made by the project activities and to look forward to the next phase of the project.

---

[4] Slides are available at: http://public.eu-egee.org/conferences/kickoff/programme/slides-dc.ppt

[5] Event's information is available at: http://public.eu-egee.org/concertation/

Action: Donatella Castelli informed EGEE participants about the current status and plans of the DILIGENT project[6]. In particular, she described the plans for adoption and exploitation of gLite middleware. Seven other members of the DILIGENT team attended the conference sessions and had both formal and informal discussions with key EGEE technical representatives. A dissemination stand was set up by DILIGENT dissemination team.

Result: This conference offered the opportunity to exchange information on the current status and plans of the DILIGENT and EGEE projects.

2.2.2   Training

DILIGENT has been in contact with EGEE NA3 team in order to organise specific training events in collaboration with DILIGENT partners. EGEE NA3 group "User Training and Induction" main goal is to produce a portfolio of training material and courses and to use this material to train a wide variety of users both internal to the EGEE consortium and from external user groups.

A first training event was organised jointly by DILIGENT, EGEE and DELOS (The Network of Excellence on Digital Libraries) and was held last April in Athens. The main purpose of this event was to foster a better cross-fertilization between the DLs and Grid domains involving representatives from both communities.

"Grid Technologies for Digital Libraries" (#20), Athens, 16th April 2005

Purpose: This tutorial, addressed to a heterogeneous audience coming from both DLs and Grid areas, introduced the DLs future directions and technology expectations, and the current available Grid technology that can be exploited to meet the requirements of the new DLs.

Abstract: Digital libraries are now moving far beyond any connotation of the term "library", and are rapidly shifting towards more general systems, sometimes known as Dynamic Universal Knowledge Environments. A major advantage of these environments is that they will permit groups of individuals, collaborating together to achieve a common goal, to access, discuss and enhance on-line shared complex, multimedia, multitype digital information objects. A radical change in current digital library systems is needed if these new collaborative environments are to be provided. New technologies must be investigated to support their implementation and new organisational and development models must be introduced. By providing a framework that supports a highly coordinated sharing of resources where large computational and storage capabilities can be exploited, the Grid is one of the most interesting candidate basic technologies for building these new generation digital library systems.

Participation: The tutorial was attended by 28 people coming from 9 different countries.

Result: This first experiment of joint event has demonstrated its relevance. Many commonalities arose during the presentations made by key players of the two areas and

---

[6] The presentation is available at:
http://indico.cern.ch/sessionDisplay.py?sessionId=7& slotId=0&confId=0513#2005-04-18

interesting stimuli for further discussion emerged. Plans for the organisation of future events of this kind, as well as technical trainings for Grid application developers and infrastructure administrators are under discussion.

2.2.3   Applications Support

DILIGENT, as one of the applications relying on EGEE middleware, is working in collaboration with EGEE NA4 group. The EGEE NA4 group "Application Identification and Support" provides support for applications: High Energy Physics, Biomedical and Generic applications.

DILIGENT plans to incrementally provide new middleware requirements (interaction #6) though a common template and official procedure proposed to and adopted by EGEE user communities. The requirements are submitted though a web interface[7] to a requirement database to be then examined by the EGEE Project Technical Forum (PTF) for prioritization and implementation.

DILIGENT representatives contributed to provide feedback to EGEE networking activities during the EGEE 2nd Project Conference. DILIGENT is maintaining contacts with EGEE NA4 representatives and UIG (interactions #26, #11) in order to provide them further feedback on user needs and cooperate with them putting in place an effective support strategy for new EGEE user communities.

DILIGENT is also contributing to EGEE NA4 Wiki site (interaction #15) which purpose is to promote and support exchange between EGEE user communities.

2.3   Interactions related to Operations

The interaction between EGEE and DILIGENT took place in various forms and at occasions, as exposed in the previous chapters. Therefore, the future activities of DILIGENT have been planned on the basis of the knowledge and experiences achieved in order to ensure successful deployment and adoption of EGEE middleware by DL applications. It was decided that the DILIGENT consortium starts deploying DILIGENT Grid infrastructures, based on gLite, for developments and testing in May 2005, just after the release of gLite 1. The following sections provide the details of this planning.

2.3.1   Background

Investigation on the DILIGENT infrastructure(s) deployment started in the December meeting at CERN with EGEE testing/integration teams (interaction #13). At that meeting, DILIGENT partners got a first idea on the main guidelines that must be followed to successfully deploy a working infrastructure. These guidelines are summarised as follows:

- one testbed should have at least three sites;
- each site should have at least three machines;
- each machine should have one (or a small number) of gLite components installed;
- infrastructures for development and for testing should be independent;
- attention should be given to network constraints (firewalls, bandwidth, etc).

---

[7] Old database: http://egee-na4.ct.infn.it/requirements/
New database: https://savannah.cern.ch/support/?group=egeeptf

Another useful information given at the meeting by the testing and integration teams was the description of their infrastructures. This concrete example of a working deployed infrastructure, together with the above guidelines and the gLite study performed up to end of January, provided to the DILIGENT partners the base to start planning and defining the DILIGENT infrastructures. An example of a possible deployment scenario is presented in Figure 4 (EGEE testing infrastructure as it was on February 2005).
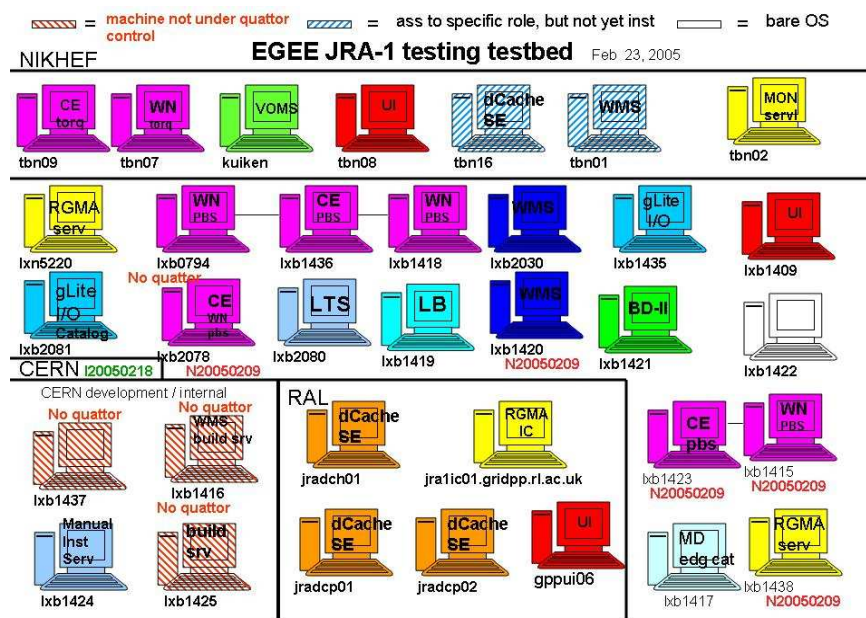

Figure 4: EGEE testing testbed on 23/02/05

2.3.2 DILIGENT Infrastructures

The DILIGENT consortium decided to setup a DILIGENT development infrastructure spanning across different sites, with partners contributing hardware for the purpose of setting up a gLite based Grid. This infrastructure will be used for the development activity. Deployment of gLite middleware started at the beginning of May 2005.

Partner	No. of Machines
CNR-ISTI	7
FhG	2(+2)*
UoA	10(+5)*
Fast	*(to be decided)*
UMIT (+ETH)	2(+3)*
Eng	3

Table 8: Resources contribution to DILIGENT development infrastructure.

(*) additional resources

gLite provides a service oriented architecture for various functionalities, and these services can be deployed on different nodes of a Grid. DILIGENT partners will focus on different aspects of building up a Grid based digital library system. Keeping this in mind, it was

decided that for the development infrastructure described above, the partners should particularly concentrate on the setup of those services which are required for the functionalities they are responsible for. Table 9 summarizes this assignment.

Partner	Functionality	gLite Components
CNR-ISTI	*Information Services*	R-GMA
FhG	*Data Management*	SE, LC, FPS
UoA	*Data Management, Job Management*	SE, LC, FPS
Fast	(*to be decided*)	
UMIT (+ETH)	*Job Management, Data Management*	SE, LC, FPS
Eng	*Security*	VOMS

Table 9: Responsibilities for system functionalities and associated gLite components.

Additionally, the setup of a separate DILIGENT testing infrastructure starting from July 2005 is also planned. Planned resources are listed in Table 10..

Name of Partner	No. of Machines
CNR-ISTI	2
Fast	(*to be decided*)
4Dsoft	3
Eng	3

Table 10: DILIGENT partners' resources contribution to DILIGENT testing infrastructure.

2.3.3   Foreseen interactions

The deployment of gLite in the DILIGENT development and testing infrastructures will surely offer the opportunity to further test and provide feedback on EGEE installation procedures and documentation. This phase is not expected to be easy, but it is expected to provide opportunities for further learning for DILIGENT partners on one side and progress in middleware setup and support for EGEE, on the other side.

Considering the lack of experience of DILIGENT partners on gLite deployment and use, all occasions for improving and accelerating the learning process will be considered and exploited, such us:

- participation to a focussed tutorial on deployment;
- participation to related open discussion forums;
- collaboration with EGEE testing team on installation procedures.

This experience could also be the opportunity to discover and formulate further requirements to EGEE.

2.4 Feedback to EGEE

At the time of writing of this first version of the deliverable, most of the interaction between EGEE and DILIGENT was devoted to anticipate as much as possible the understanding of the different aspects of the gLite middleware. gLite has been released only recently and the design of the DILIGENT services just started, consequently the feedback activity produced for the time being only a number of observations and expectations that concerns the role of DILIGENT as of a complex, distributed, dynamic service-oriented application. These have been expressed during the joint meetings, and partially submitted through the EGEE official communication channels with the user communities.

2.4.1 Mechanisms

Feedback has been provided to EGEE using several mechanisms and communication channels:

- **EGEE requirements database** – the EGEE Project Technical Forum (PTF) involves representatives from all activities and is intended to be a forum where a consensus on technical issues can be forged. One of the tasks of the PTF is to track the EGEE project requirements. Requirements are submitted by the user communities to the requirements database through a designated administrator. PTF members will then analyse, accept/refuse and rank the requirements.

  DILIGENT requirements will be discussed and ranked within DILIGENT consortium. The authorised representative will check to see if there is any duplication of the proposed new requirements with those that already exist. Then she/he will add the new requirements or update the existing ones.

- **Mailing lists** – a discussion mailing list glite-discuss@cern.ch has been set up after the technical meeting #23. It will be used to discuss on experiences, problems and best practices with the gLite middleware.

- **Web site** – a public Website[8] already exists where the DILIGENT partners report the results of the tests on the gLite middleware and on the DILIGENT infrastructure. Also the NA4 Wiki site[9] will be used to exchange with other user communities.

- **Technical meetings** – the series of joint technical meetings will continue.

- **Documentation** – in addition to the deliverables produced by the two projects, documents created as inputs for the technical meetings and minutes of them will be circulated among the interested participants of the two projects. References to the published material produced will be maintained in Deliverable 3.1.1, which will also summarize the main results of the feedback.

- **Training** – coordinated training events will be organised. They will be opportunities to present the plans for the DILIGENT services as soon as their design will be consolidated. It is expected that the services of the Collective layer, like those dedicated to the workflow management, metadata management and search management, will provide useful insights also on the design of gLite middleware services at lower level.

---

[8] DILIGENT gLite experimentation web pages: http://diligentproject.org/content/section/7/129/

[9] EGEE Generic Applications wiki site: http://egee-na4.ct.infn.it/genapps/wiki/

2.4.2   Early feedback

The requirements below, published on the EGEE Project Technical Forum database, illustrate an example of the type of generic feedback produced by DILIGENT so far and shared with other gLite user communities.

- The application programming interface to middleware services should be available in C++ and Java. It installation should be lightweight. It should cover all services accessible from the user interface host and the worker node: jobs submission, data manager, information system, etc.

- Anonymous access to restricted services is required for users accessing the Grid through specialised Grid portals.

- Parallel execution is mandatory for some applications. The middleware is expected to enable parallel job submission (by specifying a number of hosts to allocate at submission time, at least at a site scale) and provide a message passing interface (preferably MPI).

- The middleware should permit the execution of a large numbers (thousands) or short jobs (minutes at most) without introducing a prohibitive pay-off

- It should be possible for a restricted category of users (e.g. surgeons) to order high priority jobs that will execute immediately, pre-empting resources if needed.

- It should be possible to execute jobs on an input datasets (i.e. one job repeated as many time as the number of input files).

- It should be possible to execute compound (or pipelined) jobs: jobs composed of multiple unitary tasks with any directed graph of execution flow. The mechanism should allow translation of output data sets into input data set of consecutive tasks. The mechanism is expected to handle full input datasets.

- It should be possible to execute interactive jobs (jobs with a communication between the execution host and the user interface). The communication may be shell-based or application-specific (it should be possible to open a socket to transfer interactive feedback according to the application protocol). Resources reservation may be needed to ensure that interactive jobs are started at a precise time (the user has to be available when the interactive application starts).

- It should be possible to specify data required by a job. The job submission mechanism should ensure that the data is accessible without further work once the job is started (automatic data replication, etc).

- It should be possible to make advance reservation of resources and to update such reservation prior to job execution.

- It should be possible for a user to list its running jobs. For each job it should be possible to be notified of the job status progress through an API to enable jobs monitoring by an application.

- It should be possible to encrypt data on disk to prevent data leaks at the storage site level.

- It should be possible to implement an interface to new storage system to make them interoperable with the Grid middleware. This implies the availability of a standard Grid storage interface with flexibility to endorse various access control mechanisms

- Creation of new VO and propagation of VO rights to existing sites should be a simple process to ease new applications integration.

- Users will need to belong to several groups (a user is likely to take part in several projects and full VO partitioning of the user right and files is not convenient). Therefore it is expected that the user can register to several subgroups inside the VO or that several VO can be created and that a user can belong to several of them. Each user

group should have the possibility to choose whether or not to share resources with the other VO groups. It is unclear whether the VOMS already provide a support for multiple VOs registration.

More general comments and suggestions are provided here:

- All EGEE user communities, as well as many Grid related initiatives, are in close contact with EGEE. EGEE has established mechanisms for supporting at best their work and be informed of the status of their progresses. At the same time EGEE user communities and related projects are not necessarily aware of each other's activities and objectives. At present there is not clear direct channel of information and communication between them. For instance, information on which other users are experimenting on gLite and what they are doing, which organisations are active in a specific domain (e.g. DLs, search engines, etc.) or in a specific geographical area, are generally provided through personal contacts with EGEE representatives. Therefore, there is a need for improving the communication and coordination between these communities at different levels (dissemination, collaboration, technical exchanges) and not necessarily rely on EGEE informal communication to ensure these contacts. This could also help the transmission of knowledge from Grid experienced users to "newcomers". For instance, representatives from all EGEE user communities attend the EGEE conferences. Therefore these conferences could become also opportunities for setting up specific sessions where the users can share and discuss on they experiences and plans.

- New EGEE user communities often feel lost when they first look at EGEE activities and documentation. Without receiving active support from experienced persons they do not know where to find relevant information or simply where to start looking in relation to their level of understanding and their objectives. EGEE is already investigating the necessity of providing a unique user-oriented interface to the project's documentation. This will be of great help for the users. New users and new projects often ask the same questions when they start working with EGEE, only afterward they differentiate in more specific needs and requirements. It would be already very useful to have a Web page for EGEE related projects, like DILIGENT, which could provide all the information and guidelines they require at their start up. See section 1.1 of this document for a detailed list of needs and possible actions that have been identified in DILIGENT.

- It seems there is a lack of documentation for communities, like DILIGENT, who would like to set up their own development infrastructure based on gLite to conduct initial tests and experimentation on the middleware installation and usage. Having clear deployment instructions such as hardware requirements, distribution of services between sites, suggested configurations or examples of possible deployments with variants depending on the user needs, as well as specialised trainings, would provide an extremely useful guideline.

2.4.3   Plans for the near future

In the immediate future the main opportunities for providing feedback on the adoption of gLite in DILIGENT will be:

- the planned deployment of the DILIGENT infrastructures (described in Section 4.2);
- the gLite experimentation (described in Section 2.3), which will be extended to involve all the partners;
- the design of the DILIGENT specific services, each having its own requirements on the exploitation of the gLite middleware;
- the definition and realization of a simple DL application extracted from the ImpECt scenario. This will involve the implementation of a simplified version of the services that are needed to support the chosen scenario.

DILIGENT will be one of the first communities, outside the EGEE consortium, to operate such intensive experimentation on gLite, which has started about one month after the publication of the first release of gLite. It is expected that the set up of the DILIGENT gLite-based infrastructures will initially require a large effort. The experimentation, and the realization of the application scenario, will be conducted in parallel by all the DILIGENT partners, which belong to different types of organisations (i.e. industries and research organisations) with different expertise. Each partner will focus its attention on the specific range of functionalities that are needed to support the services that it has to develop. Given the novelty of the experimentation, the variety of the tests and the heterogeneity of the testers, it is expected that very useful outcomes will be produced.

In parallel to the feedback provided by the experimentation, the detailed design of the DILIGENT infrastructure services will clarify how DL services will build on top of gLite and it will highlight to what extent the functionalities provided by this middleware are sufficient to support the DL-like applications.

In addition to the above activities, the DILIGENT partners also plan to intensify their contribution through the EGEE established feedback channels. In particular, the project intends to:

- identify and implement a number of test cases of interest to DILIGENT, in coordination with EGEE, as part of the gLite middleware testing and experimentation activity;
- submit of DILIGENT specific requirements to the EGEE middleware requirements database;
- exchange with other user communities adopting gLite though EGEE NA4 activities.

2.4.4   Open issues and difficulties

As early adopters of gLite and mostly new to the Grid environment, the DILIGENT partners are now facing with a number of difficulties that are common to other EGEE user communities. Some of these difficulties are clearly reported below since making them explicit may help EGEE in better focussing the user communities support activities.

- Most of DILIGENT partners had never developed an application on Grid before. In order to be able to design this application the project had to change our perspective from distributed-applications to Grid-applications and become accustomed to work in a collaborative and distributed environment where standards are not defined, policies have to be agreed and conditions can change, where the infrastructure is not dedicated but shared between different administrative domains and for different domains of application. This change of perspective has required a lot of effort, much higher than initially planned, and it has imposed a great pressure on our activities.
- DILIGENT is designing an application on a new and not yet consolidated technology which will certainly evolve quite a lot in the future. Learning and keeping us updated on the new releases of this software is not simple and it is certainly very time consuming. This difficulty is amplified by the fact that the project intends not only to exploit the gLite functionality but also reuse and expand part of the developed software. By working at "low-level" it is required to be aware of many details of the software and it is needed to keep track of any change done at that level.
- At the time of the writing of this report, the gLite design team is working at establishing the priorities among the list of still pending functionalities that will be developed before the end of the first phase of the EGEE project. The requirements expressed by DILIGENT, as those stated by the other user communities, are important input for this process. By expressing these requirements DILIGENT may increase the chances to be able to exploit soon a version of the gLite middleware that provide the functionality that

is needed by its services. Unfortunately, however, the timing of the two projects is not completely aligned. As it has been explained above, the DILIGENT project is now at the beginning of its experimentation and design activity, while the plans for development in EGEE phase one must be completed in a couple of months. In order to profit as much as possible of the opportunity given by the process of establishing priorities, the project will plan in the next months to intensify its experimentation activity and the interactions with gLite key members.

2.5   Final Considerations

During these first months of the project it has become more and more evident that close interactions between the two interrelated projects are necessary. In the case of DILIGENT, being an early adopter of EGEE new Grid middleware, exchange, coordination and dissemination of plans and results are vital activities. These allow to minimize the project's risks related to the adoption of the new technology and to ensure a successful collaboration. Being so important, there is a need for DILIGENT to dedicate a complete work package to this interaction activity and inter-project coordination. At the same time, EGEE showed to be extremely open in taking into account users' needs, requirements and feedback.

Interactions between DILIGENT and EGEE, motivations and outcome have been discussed in this document, where also early feedback has been collected. The activities of WP3.1 will continue supporting the collaboration between the two projects and they will evolve as DILIGENT technical work progresses. Further feedback to EGEE and the DL communities will be collected and reported.

3   DESIGN AND EXPERIMENTATION (M9 – M22)

This chapter describes in detail the interactions made between project month 9 and project month 22: from May 2005 to June 2006. Main motivations for interaction between the DILIGENT and the EGEE project were, during this second period, the need to keep the technical plans of the two projects aligned as well as to provide feedback to EGEE on DILIGENT experience with the gLite middleware and requirements for future enhancements.

3.1   Interactions related to Middleware

3.1.1   Joint Technical meetings

Whereas the first phase of the EGEE-DILIGENT interaction concentrated on the study of the gLite middleware through the inspection of the EGEE documentation, the second phase can be described by the concrete experimentation activities with this middleware and the formulation of the requirements. The results of these activities were discussed during a number of phone conferences with EGEE experts and contributed to the constitution of specific working groups. These working groups are studying the mechanisms of how to satisfy EGEE user needs and the priorities for future development and support of functionality provided by gLite middleware. DILIGENT experts participate in several of these working groups to contribute to the discussion of DILIGENT requirements and directly benefit from the expertise of the EGEE members.

3.1.1.1   Middleware

DILIGENT actively participated in the technical groups that are working to improve the usability and usefulness of the gLite middleware. In particular, DILIGENT joined the EGEE *"Middleware Security Group"* and the *"Joint Security Policy Group"* to discuss the authentication, authorization, and delegation of users and services in the trusted Grid enabled infrastructure; the *"Short Deadline Job Group"* to participate in the tuning of the gLite middleware in the scenarios where a huge number of short jobs must be executed; and the *"Medical Data Management Group"* in order to contribute to and learn about data security and preservation in multi-institutional storage infrastructure.

Meeting with EGEE Technical Director (#39), Geneva, November 2005

Topic/Purpose: Discuss the goal of the Technical Coordination Group (TCG) and DILIGENT participation in this group.

Participants: Erwin Laure (EGEE), Pedro Andrade and Florida Estrella (CERN, DILIGENT)

Results: E. Laure informed DILIGENT about the objectives of the new Technical Coordination Group and about how DILIGENT could be represented in this group. More information about the EGEE TCG can be found at the following link: http://egee-intranet.web.cern.ch/egee-intranet/NA1/TCG/tcg.htm.

EGEE Middleware Security Group Meeting (#42), Amsterdam, 14-15 December 2005 and EGEE Joint Security Policy Group Meeting (#43), Amsterdam, 16 December 2005

Topic/Purpose: Update on current global security architecture work. Discuss future global security architecture work.

Participants: Paolo Roccetti (ENG, DILIGENT), Andrea Manieri (ENG, DILIGENT), members of EGEE, GRIDPP, OSG, DEISA, SWITCH/EGEE-II, SEEGRID, and CESNET projects. The complete list of participants is available at:

http://agenda.cern.ch/askArchive.php?base=agenda&categ=a058129&id=a058129/attendees

Result: The meeting offered the possibility to present the DILIGENT authentication and authorization framework and to discuss about the following issues:

- Detecting ill-configured nodes
- GUIDs
- Delegation Interface
- Glexec on worker nodes
- CA namespace constraints implementations/SwissSign
- Rare RDN components
- 1SCP
- Data management security model

The minutes of the meeting are available at the following URL:

http://agenda.cern.ch/askArchive.php?base=agenda&categ=a058129&id=a058129s8%2Fminutes%2FMWSG7-openissues-conclusion.pdf

Documentation: The GGF CAOPS draft document on namespace constraints provides, with the additional/modifications mentioned in the open issues-conclusion presentation, the root that will be implemented as stated by the relevant M/W providers present. It is accessible from:

http://agenda.cern.ch/askArchive.php?base=agenda&categ=a058129&id=a058129s8%2Fdocument%2Fdraft-ggf-namespace-policy-20050929-3.pdf

Meeting with EGEE NA4 (#45), Phone Conference, December 2005

Topic/Purpose: Meeting with NA4 people to discuss the DILIGENT requirements and the DILIGENT involvement in TCG workgroups - MDM and SDJ.

Participants: Charles Loomis (EGEE), Birger Koblitz (EGEE), Pedro Andrade (CERN, DILIGENT), Pasquale Pagano (CNR-ISTI, DILIGENT)

Result: This first meeting between the representatives of the two projects offered the opportunity to start the discussion on specific DILIGENT technical requirements. In particular, this meeting contributed to the identification of the technical working group the DILIGENT project should participate in by providing a framework for presenting the DILIGENT requirements related to the Grid middleware,.

Documentation: The following documents have been produced and can be accessed from:
- list of the requirements -
https://uimon.cern.ch/twiki/bin/view/DILIGENT/DiligentRequirements;

- textual description of these requirements and of the DILIGENT vision –
https://uimon.cern.ch/twiki/pub/DILIGENT/DiligentRequirements/DiligentVision.doc

Meeting with EGEE NA4 (#48), Phone Conference, January 2005

Topic/Purpose: DILIGENT joined the MDM and SDJ workgroups. Discussion with Cal Loomis about DILIGENT participation in the EGEE TCG JP workgroup.

Participants: Charles Loomis (EGEE), Birger Koblitz (EGEE), Pedro Andrade (CERN, DILIGENT), Pasquale Pagano (CNR-ISTI, DILIGENT)

EGEE TCG JP workgroup meeting (#49), Geneva, 28th February

Topic/Purpose: Kick-off meeting of the TCG JP workgroup

Participants: ATLAS, LHCb, CMS, and SA1 members, Pedro Andrade (CERN, DILIGENT)

Agenda:
- Workgroup presentation
- Projects requirements and expectations
- GPBox solution
- Site Configuration solution

Results:
- Technical discussions should continue in the workgroup mailing list and from time to time face-to-face meetings will be set up
- Each user groups should prepare a short requirements document
- G-PBox should be analyzed as a possible solution. Discussion about it will continue in the mailing list
- Simple site configuration can also be seen as a partial solution. It is evident that such site configuration will not satisfy all the requirements. Nevertheless, it can be important as a short-term solution. SA1 will circulate a possible an example of a site configuration

EGEE TCG SDJ workgroup meeting (#52), March 2006

Topic/Purpose: Update on current Short Deadline Job (SDJ) work. Discuss future activities.

Participants: Cecile Germain-Renaud (WG Leader), Charles Loomis, Massimo La Manna, Fabrizio Pacini, Dieter Kranzlmueller, Martin Polak, Christoph Langguth (UNIBAS, DILIGENT), Laura Voicu (UNIBAS, DILIGENT).

Agenda: The discussion covered the following topics:
- presentation and discussion of involved projects and use cases
- agree on a common definition of what exactly SDJs are
- report on initial experiments with specially configured SDJ CEs

- discussion on required changes to the underlying gLite middleware to better support SDJs
- presentation and discussion of gLogin
- preparation of the SDJ report (structure, contributions, ...)

Results: Discussion on the definition of what is a short deadline job, which are the most typical use cases that involve short deadline jobs, and the set of requirements needed to assure the requested QoS (Quality of Services). A valuable solution was also suggested. This solution is based on a specific configuration of Torque LRMS and on a modified version of WMS that will be provided with gLite 3.2. A detailed description of this activity is reported in the following document:

http://egee-intranet.web.cern.ch/egee-intranet/NA1/TCG/wgs/SDJ-WG-TEC-v1.1.pdf

Middleware Security Group Meeting (#53), Geneva CERN, 7-8 March 2006

Topic/Purpose: Update on current global security architecture work. Discuss future global security architecture work.

Participants: Paolo Roccetti (ENG, DILIGENT), members of EGEE and OSG projects. The complete list of attendees is available at:

http://agenda.cern.ch/askArchive.php?base=agenda&categ=a061444&id=a061444/attendees

Agenda:
- EGEE Security Architecture Assessment
- Follow up of GF16 Authorization interoperability meeting
- New convention on VO names
- G-PBox performance benchmark
- Delegation update

The detailed agenda of the meeting is available at:
http://agenda.cern.ch/fullAgenda.php?ida=a061444

Results: An improved understanding of VOMS model and G-PBox capabilities has been achieved between attendees of the meeting. Technical conclusions are reported at the following address:

http://agenda.cern.ch/askArchive.php?base=agenda&categ=a061444&id=a061444%2Fminutes%2FMWSG-8-Conclusions.txt

Meeting with EGEE technical coordination (#55), Geneva, 3rd May 2006

Topic/Purpose: Meeting with EGEE Technical Director Erwin Laure and EGEE JRA1 Activity leader Claudio Grandi to present DILIGENT usage of gLite and discussion on major requirements.

Participants: Erwin Laure (EGEE), Claudio Grandi (EGEE), Roberta Faggian Marque (CERN/DILIGENT), Pedro Andrade (CERN/DILIGENT)

Results:

- DILIGENT has no problems with scratch installs. Only test data is used therefore no need exists to preserve data through the upgrade. An update procedure from the currently deployed gLite 1.4.1 to gLite 3.0 is not needed. The only exception is VOMS.
- EGEE needs to checked if VOMS from gLite 1.4.1 can be upgraded to VOMS from gLite 3.0.
- DILIGENT has adopted the new data management components of gLite 3.0, however, there is a critical need of web service interfaces to them.
- DILIGENT relies on R-GMA as the information system for the gLite resources. The implications of this usage in production and in tools like LCG-utils should be analyzed.

EGEE TCG JP meeting (#56), Geneva, 4th May 2006

Topic/Purpose: 3rd meeting of the EGEE TCG JP workgroup

Participants: ATLAS, CMS, EGEE SA1 and DILIGENT

Agenda:

- Short term goals
- Report from NIKHEF
- Report from CNAF
- Report from INFN
- Long term goals
- Testbed to use

Results:

- Two solutions to address the current requirements of were identified:
  - One solution is based on a static site configuration and the use of VOViews to publish information of shares or priorities associated to a VOMS group or role. Using small modifications the Resource Broker can take this additional information into account and direct jobs based on VOMS group or role of the user certificate. The simplicity of this solution is based on the same mechanism that allows several VOs share a CE. The disadvantage is that configurations are basically static, that in practice only a limited number of groups could be supported and information on these groups is published. The main requirement is to have GLUE 1.2.
  - The second solution is based on the G-PBOX. The mechanism has many advantages, ranging from a dynamic configuration mechanism to the support of large number of groups etc. There was general agreement that G-PBOX is the solution in the medium term. The software is already included in gLite, but has not yet passed certification and deployment.
- The testbed here these solutions should be apply was also discussed:
  - The use of the EGEE PPS for these scenarios does not make too much sense since larger installations are necessary to study the reactions of the fair share

mechanism. Also important ingredients are not really middleware, but rather site configurations.

- o It was generally agreed to move immediately to the production sites. The necessary work is equivalent to the normal tuning of batch systems in computer centres and does not imply a big risk for the service.
- o It was discussed to have possible setups in NIKHEF, RAL, and CNAF. At this point in time it was not considered necessary to find further centres, since it can be assumed that after the first steps have been taken, other centres will follow in a natural way.

3.1.1.2 Integration

DILIGENT's first software release (Release Alfa) is expected to be available by project month 24, this means August 2006. Consequently, the software integration activities - WP1.8 Integration and Deployment - started its work in project month 19.

EGEE, as one of the biggest Grid projects in Europe, has a very complex and organised integration system. gLite developers are spread over many countries, and therefore EGEE was obliged to define a clear process to assemble its software and to produce a unique release. Although on a smaller scale, DILIGENT software is also a product of distributed development. Seven partners are contributing to the development of the DILIGENT software. Therefore, the DILIGENT integration activity has followed some of the EGEE guidelines and practices. The study of EGEE practices has started form the analysis of the available documentation related to integration procedures and was refined in a number of meetings organised with EGEE JRA1 integration team leader and ETICS technical coordinator (#33 and #50). ETICS is an EU project that plans to deliver a certification process for Grid and distributed software by offering build and test services via secure web interfaces and command line tools. The information exchanged during these meetings was extremely helpful to better understand the EGEE software release cycle, release types, integration infrastructures among other best practices.

Integration and Deployment (#33), Geneva, 7th July 2005

Topic/Purpose: Understand the organisation and functioning of the EGEE infrastructures and integration activities and what can be applied or adopted in DILIGENT

Participants: Pedro Andrade, Roberta Faggian Marque, Florida Estrella Cainglet, Alberto Di Meglio (EGEE JRA1 integration team leader)

Results:

- A better understanding of the EGEE integration model was reached
- A proposal of the DILIGENT infrastructures and the DILIGENT integration cycle was defined. This proposal is depicted below:

Figure 5: DILIGENT infrastructures
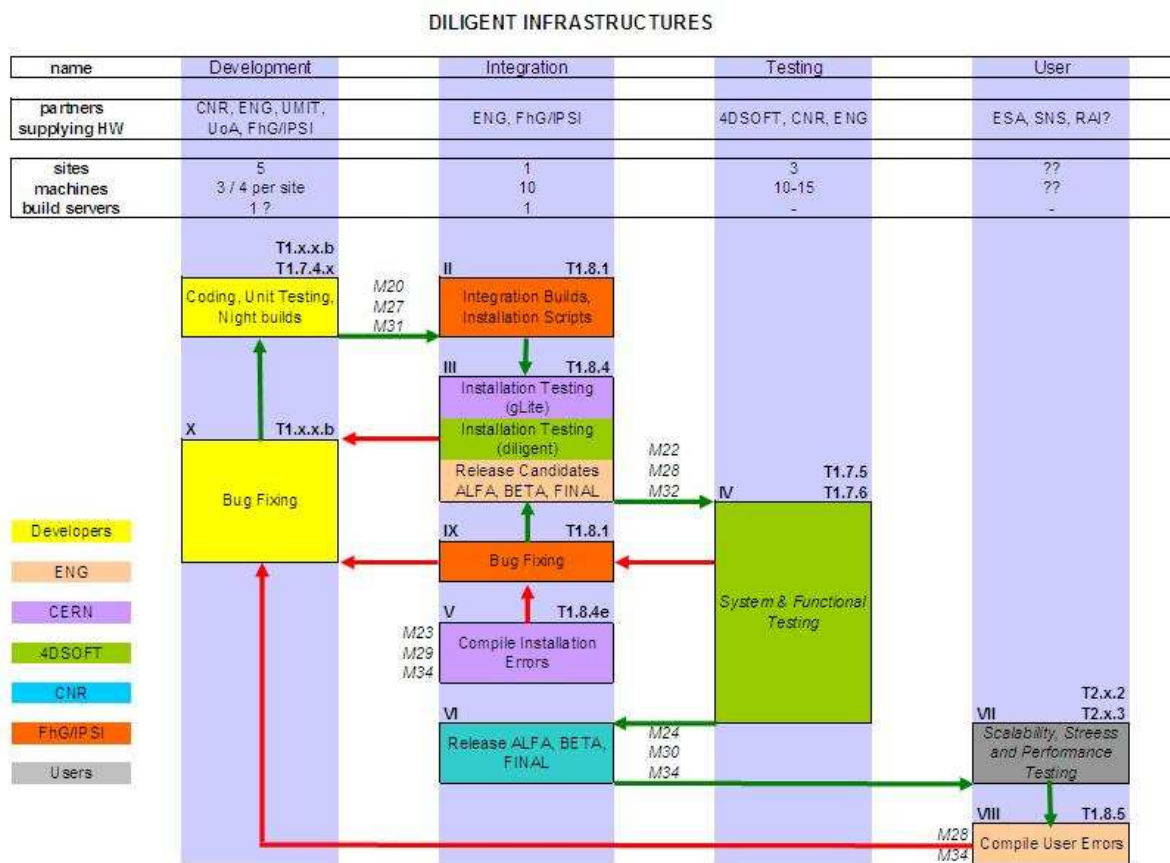
Integration and ETICS (#50), Geneva, 19th February 2006

Topic/Purpose: Discuss about DILIGENT integration activity and ETICS adoption

Participants: Pedro Andrade, Roberta Faggian Marque, Alberto Di Meglio (EGEE JRA1 integration team leader and ETICS coordinator), Marc-Elian Begin (ETICS WP3 leader)

Agenda:
- ETICS presentation - ETICS usage in DILIGENT
- Discussion about the DILIGENT integration activity

Results:
- ETICS can provide some of the tools that DILIGENT needs for its integration activity:
  - build framework
  - automatic distributed execution of deployment tests
  - automatic deployment (planned for the end of 2006)
  - build and test reports
- To simplify the integration activities gLite and DILIGENT services could run on two different sets of machines. Possible dependencies to be clarified with the DILIGENT partners.

- The gLite services, already installed in the DILIGENT testing infrastructure, could be shared by the integration and testing activities. Possible incompatibilities to be clarified with the DILIGENT partners.

3.1.1.3   Testing

Active since the beginning of the project, the DILIGENT testing activity is responsible for the definition of the DILIGENT test strategy.

As in the integration activity case, the DILIGENT testing team can also learn and benefit from the EGEE experience. To promote this knowledge exchange, one DILIGENT member from the DILIGENT partner coordinating the testing activity (4DSOFT) spent one week at CERN working with the EGEE JRA1 testing team (#30). This allowed him, and hence DILIGENT, to have a closer look on what EGEE is doing with respect to testing and to understand how that can be applied in DILIGENT.

3.1.2   gLite experimentation

The experimentation with the gLite software is an important activity since it provides the basis for the design and implementation of the DILIGENT services. Whereas in the first phase of the project the effort was focused in getting to know the middleware, in this second phase the need was in the analysis of the gLite functionalities and in the understanding of how those functionalities can be exploited by the DILIGENT services.

From project month 9 to project month 22 three major experimentation activities were initiated and completed:

- gLite experimentation with DILIGENT use cases
    - Period: from June 2005 to August 2005
    - Objective: implementation of test cases focused in the interaction between gLite and some DILIGENT services
- Distributed gLite experimentation with DILIGENT use cases
    - Period: from September 2005 to November 2005
    - Objective: implementation of distributed test cases focused in the interaction between gLite and some DILIGENT services
- gLite 3.0 experimentation
    - Period: from February 2006 to March 2006
    - Objective: study and experiment the new services included in gLite3.0

These three experimentations are described in detail below.

3.1.2.1   gLite experimentation with DILIGENT use cases

To better understand how gLite can be exploited by the DILIGENT DL Creation and Management services and by the Process Management services, CERN carried out a set of tests with the following goals:

- implement test cases focused in the interaction between gLite and the DIS and Keeper services, as described in the D1.2.1

- implement test cases focused in the interaction between gLite and the Process Management services, as described in the D1.5.1
- gain more experience with gLite in order to better support the other DILIGENT partners
- provide material for the first implementation of the DILIGENT services
- provide the necessary Grid functionalities for the ESA experimentation scenario

The tests executed in this experimentation were grouped in 3 areas:
- Information System Tests – interaction between the DIS service and gLite
- Data Management Tests – interaction between the Keeper service and gLite
- Job Management Tests – interaction between Process Management and gLite

All tests were executed in the DILIGENT development infrastructure. If needed, because of malfunctioning or unavailability of the DILIGENT machines, the EGEE prototype infrastructure was used in alternative.

Finally, it is worth noting that for most of the tests executed in this experimentation the results were generated using the TestManager tool (#31):

http://meb.home.cern.ch/meb/Software/TestManager.htm

TestManager, used by the EGEE testing team, is a tool that allows the execution of tests and generates an integrated HTML report. The tests can be written in an xUnit framework (CppUnit, JUnit, PyUnit). The common usage of this tool is a good example of the tight collaboration between DILIGENT and EGEE.

JIRE application & Reuters corpus

In most of the following tests, one DL specific application called JIRE and the data from Reuters were used.

The JIRE application is a set of JAVA classes that processes a given number of files, trying to extract information about its content. The JIRE application receives as input one directory containing the files to process. JIRE processes all files within this input directory and all files of all subdirectories one level below. The part of the Reuters corpus used was composed by 800.000 files. These files were grouped in 13 directories organised by month.

Information System Tests

Test 1

Information consumer that extracts information about services and sites.
The source code can be found here:

https://twiki.cern.ch/twiki/pub/DILIGENT/GLiteDiligentTC/DGT_RGMA_Consumer_v1.0.tgz

The results can be found here:

http://prodrigu.home.cern.ch/prodrigu/rgma_reports

Test 2

Information producer and information consumer that publishes and retrieves information about Computing Elements (CE) and Storage Elements (SE). The published information follows the GLUE schema (http://infnforge.cnaf.infn.it/projects/glueinfomodel).

The source code can be found here:

https://twiki.cern.ch/twiki/pub/DILIGENT/GLiteDiligentTC/DGT_RGMA_Producer_v1.0.tgz

The results can be found here:

http://prodrigu.home.cern.ch/prodrigu/rgma_reports

Test 3

WSRF service that consumes information from R-GMA Server. One main method (getRGMAdata) is made available and is responsible for querying the R-GMA server, retrieve the requested data and return it to the user. This method receives as input the type of data to retrieve, such as:

- published sites
- available services and their status
- computing element (CE)
- storage element (SE)

The following diagram illustrates how this service, the client and R-GMA interact:
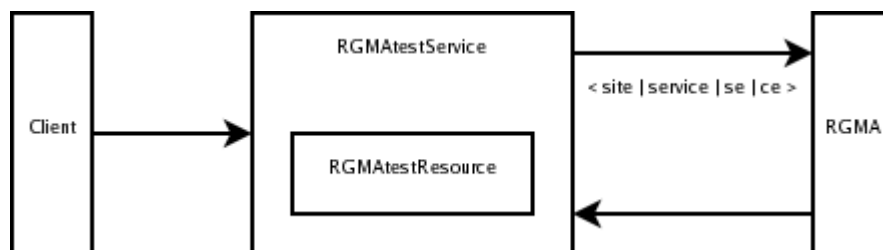

Figure 6: Components interaction

The source code can be found here:

- https://twiki.cern.ch/twiki/pub/DILIGENT/GLiteDiligentTC/DGT_RGMA_WSRFConsumerClient_v1.0.tgz
- https://twiki.cern.ch/twiki/pub/DILIGENT/GLiteDiligentTC/DGT_RGMA_WSRFConsumerService_v1.0.tgz

Data Management Tests

Test 1

Copy, register and make available in "the Grid" the Reuters corpus. The correct upload of these directories/files and their correct insertion in the gLite catalogue is verified.

The source code can be found here:

https://twiki.cern.ch/twiki/pub/DILIGENT/GLiteDiligentTC/upload_data_v2.0.sh

As a result of this test, part of the Reuters corpus was successfully uploaded to the gLite storage elements. In particular of the 13 directories that composed the Reuters corpus, two were uploaded: 1996-09 and 1998-10. The average upload time of each file was 3.8 seconds.

Test 2

WSRF services running on Java WS Core that:

- upload DILIGENT packages to one DILIGENT repository through the Package Repository Service
- deploy a DILIGENT package in one available hosting node. This is done using two services: Digital Library Manager service and Hosting Node Service.

Three services were made available:

- DLmanager service: WSRF service implementation that exposes an interface to users to deploy registered packages. The service selects an Hosting Node and sends it a request to deploy a given DILIGENT package. The user only needs to specify the package name. The service also implements a stateful resource to keep trace of the deployed packages.
- HNmanager service: WSRF service implementation that exposes an interface to DLmanager service to deploy packages on a given Hosting Node. The service accepts deploy requests from the DLmanager service, asks to the PKGrepository service to retrieve the package from the SE and then it deploys the given package in the local container. The service also implements a stateful resource to keep trace of the deployed packages on the local Hosting Node.
- PKGrepository service: WSRF service implementation that exposes methods to manage DILIGENT packages. This service is invoked to validate and register a package, uploading it in the SE. In this case, the service is invoked passing the local path of the package to upload and returns the LFN of the stored file. Is it also possible to retrieve a package passing as an argument the package name. In this case the service returns the local path of the retrieved package. The service has also a stateful resource to keep trace of the registered packages.

In addition to these services, one Java client was also implemented. It provides the basic management functionalities of the DILIGENT packages. In particular it provides methods in order to contact the PKGrepository service to validate and register packages. The application gives also to the user the possibility to contact the DLmanager service to deploy a registered package. The following diagram illustrates how these there services, the client and the SE interact with one another:
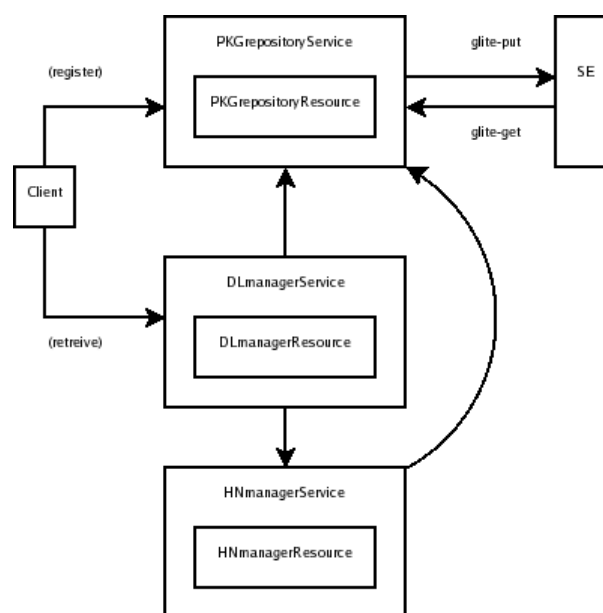

Figure 7: Components interaction

The source code can be found here:

- https://twiki.cern.ch/twiki/pub/DILIGENT/GLiteDiligentTC/DGT_DM_DLmanager_v1.0.tgz
- https://twiki.cern.ch/twiki/pub/DILIGENT/GLiteDiligentTC/DGT_DM_HNmanager_v1.0.tgz
- https://twiki.cern.ch/twiki/pub/DILIGENT/GLiteDiligentTC/DGT_DM_PKGclient_v1.0.tgz
- https://twiki.cern.ch/twiki/pub/DILIGENT/GLiteDiligentTC/DGT_DM_PKGrepository_v1.0.tgz

<u>Job Management Tests</u>

Test 1

Java application which uses the WMS Java API to execute the JIRE application. The input data was retrieved directly from the UI or from one SE. This application can be executed in 3 different ways:

- runJobManager: interacts with WMS. The following parameters are expected:
    - data_source: data source (either UI or SE)
    - submit: submit n number of jobs
        - directory: name of the directory which contains the files to analyze
        - number of files: total number of files to use
        - number of jobs: number of created jobs. Each job will use as input files: #files/#jobs.
    - status: get the status of the submitted jobs
    - status_done: get the status of the submitted jobs and wait DONE state
    - output: get the output of the submitted jobs
- testJobManagerUI: submits, waits jobs to finish and gets the output using data from UI (no parameters expected)
- testJobManagerSE: submits, waits jobs to finish and gets the output using data from SE (no parameters expected)

The source code can be found here:

https://twiki.cern.ch/twiki/pub/DILIGENT/GLiteDiligentTC/DGT_JStest_v3.0.tgz

This test was performed in 5 different Workload Management Systems: 2 from the EGEE infrastructures (prototype infrastructure and pre-production infrastructure) and 3 from the DILIGENT development infrastructure. 1000 files were processed using 10 jobs (100 files each job) and then 1000 files were processed using 100 jobs (10 files each job). The average execution time of each file was 0,61 seconds. More information about the tests and the results can be found here:

https://twiki.cern.ch/twiki/bin/view/DILIGENT/GLiteDiligentTC

3.1.2.2   Distributed gLite experimentation with DILIGENT use cases

During September and October 2005, a set of distributed and coordinated experiments were conducted between several DILIGENT partners. Each experiment tested the basic functionality of job submission and data transfer of the gLite middleware. The objectives of this experimentation were:

- check the suitability of the gLite middleware for the DILIGENT requirements
- increase the links between the sites of the DILIGENT development infrastructure

- provide concrete statistics on the status of the DILIGENT development infrastructure

The purpose of this experimentation was to demonstrate the effectiveness of the DILIGENT Grid as one of the early adopters of gLite. The following partners took part in this experimentation: CNR, FhG, UMIT, UoA and CERN. These results were also presented at the 1st DILIGENT review in November 2005.

The setup of the gLite middleware was as follows:

Partner & component	Configuration

CNR CE	`configured to use CNR WMS (pull mode)`
	`configured to use UoA WMS (pull mode)`
	`configured to use UMIT WMS (pull mode)`

UoA CE	`configured to use CNR WMS (pull mode)`
	`configured to use UoA WMS (pull mode)`
	`configured to use UMIT WMS (pull mode)`

UMIT CE	`configured to use CNR WMS (pull mode)`
	`configured to use UoA WMS (pull mode)`
	`configured to use UMIT WMS (pull mode)`

All WNs	`configure to use its site SE`

CNR SEs	`configured to use FhG Catalog`
	`containing the following Reuters corpus directories (not the files):`
	`    /Reuters.corpus/all/1996-09`
	`    /Reuters.corpus/all/1996-10`
	`    /Reuters.corpus/all/1996-11`
	`    /Reuters.corpus/all/1996-12`

FhG SE	`configured to use FhG Catalog` `containing all Reuters corpus directories (not the files)`

UoA SE	`configured to use FhG Catalog`
	`containing the following Reuters corpus directories (not the files):`
	`    /Reuters.corpus/all/1997-01`
	`    /Reuters.corpus/all/1997-02`
	`    /Reuters.corpus/all/1997-03`
	`    /Reuters.corpus/all/1997-04`

UMIT SE	`configured to use FhG Catalog`
	`containing the following Reuters corpus directories (not the files):`
	`    /Reuters.corpus/all/1997-05`
	`    /Reuters.corpus/all/1997-06`
	`    /Reuters.corpus/all/1997-07`
	`    /Reuters.corpus/all/1997-08`

The concrete experiments consisted of the following tasks

Partner	Task
CNR	`upload Reuters corpus according to:` • `upload /Reuters.corpus/all/1996-09 to CNR SE` • `upload /Reuters.corpus/all/1996-10 to CNR SE` • `upload /Reuters.corpus/all/1997-03 to UoA SE` • `upload /Reuters.corpus/all/1997-07 to UMIT SE`

	get upload time for each directories
	submit the JIRE application to CRN WMS according to:
	execute 10 and 100 jobs in CNR CE using 10000 files from 1996-09 directory
	• Repeat for 1996-10 directory, 1997-03 directory, 1997-07 directory.
	get the execution time for each directories (execution time include job submission, download of input data to the WN from the SE, job execution and upload of the results to the SE)
	transfer Reuters corpus using CNR FTS:
	directory 1996-08 from FhG SE to CNR SE
	get copy time for each directories
FhG	upload Reuters corpus according to:
	upload all Reuters corpus to FhG SE
	get upload time for each directories
	transfer Reuters corpus using FhG FTS:
	• directory 1996-12 from CNR SE to FhG SE (different then /Reuters.corpus)
	• directory 1997-04 from UoA SE to FhG SE (different then /Reuters.corpus)
	• directory 1997-08 from UMIT SE to FhG SE (different then /Reuters.corpus)
	get copy time for each directories
UoA	upload Reuters corpus according to:
	• upload /Reuters.corpus/all/1996-11 to CNR SE
	• upload /Reuters.corpus/all/1997-01 to UoA SE
	• upload /Reuters.corpus/all/1997-02 to UoA SE
	• upload /Reuters.corpus/all/1997-08 to UMIT SE
	get upload time for each directories
	submit the JIRE application to the UoA WMS according to:
	• execute 10 and 100 jobs in CNR CE using 10000 files from 1996-11 directory
	• execute 10 and 100 jobs in UoA CE using 10000 files from 1997-01 directory
	• execute 10 and 100 jobs in UoA CE using 10000 files from 1997-02 directory
	• execute 10 and 100 jobs in UMIT CE using 10000 files from 1997-08 directory
	get the execution time for each directories (execution time include job submission, download of input data to the WN from the SE, job execution and upload of the results to the SE)
	transfer Reuters corpus using UoA FTS:
	directory 1996-08 from FhG SE to UoA SE
	get copy time for each directories

In the above experiments, two mass storage systems were tested: dCache and DPM. The performed evaluation measured the upload rate, which is defined as the number of files uploaded per second. The Reuters corpus has around 780,000 files with an average size of 4.8 kb, with a small standard deviation. dCache provides a system for storing and retrieving huge amounts of data, distributed among a large number of heterogeneous server nodes, under a single virtual file system tree with a variety of standard access methods. The Disk Pool Manager (DPM) has been developed as a lightweight solution for disk storage management with no limitation on the amount of disk space that it can handle.

The tests were performed on a very basic storage system infrastructure and were tailored to measure the capability to manage a huge number of continuous storage and retrieve operations of small files. This scenario well simulates the typical operations that occur in operational digital libraries.

The figures below show the performance of these systems when subjected to data transfer tests as previously described. dCache performed quite badly, with a high failure rate (69.06 % success rate), moreover the upload rate (16.18 files/s) was higher than DPM. DPM performed more reliably (97.26% success rate) and the upload rate was almost 3 times

slower (6.1 files/s) than dCache. These numbers and comparisons were confirmed running the tests in the DILIGENT development infrastructure and the EGEE PPS. A very basic storage infrastructure together with the huge number of concurrent and continuous operations performed on small files determined the cause of the bad performance of dCache. While the experiments have not been repeated, it is expected that better results can be obtained changing the dCache configuration in order to distribute its components on multiple servers.

More details about these experiments, including source code can be found at:

https://twiki.cern.ch/twiki/bin/view/DILIGENT/GLiteDiligentCombTests
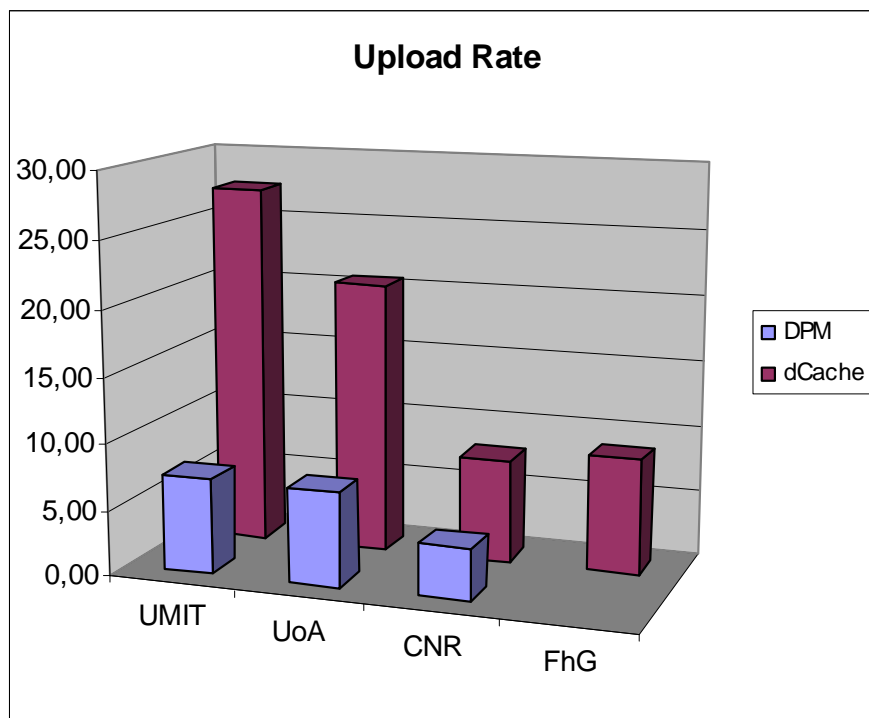
Figure 8: DILIGENT development infrastructure
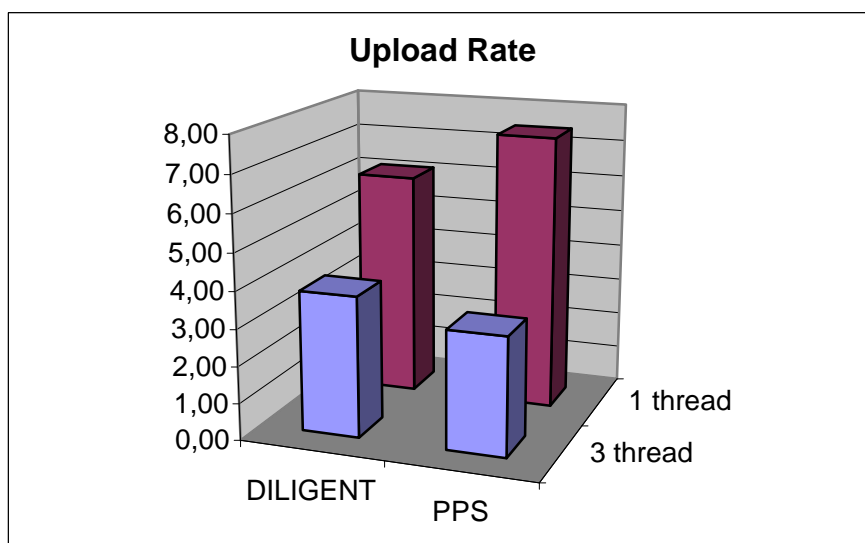
Figure 9: Upload rates for DPM

An additional experiment was performed to test the Job Submission services provided by the gLite middleware. The experiment consisted in the processing of several sets of 10000 files and 1000 files from the Reuters Corpus, splited in 1, 10 or 100 jobs. All the data to process was taken from the closest Storage Element (DPM and dCache) here the data was found.

With respect to the download of data from the SEs to the several Worker Nodes, the results again showed the maturity and higher performance of DPM with respect to dCache. As in the data upload tests, several technical problems were found with the dCache setup. This might be attributed to the faulty configuration which was determined much later in the project.

Regarding the overall distributed execution, the all process of creating several jobs, dispatch them to the several Worker Nodes and collect the results, some benefits were visible when using an higher number of jobs (higher level of parallelization) but this advantage was not obvious:

    average execution time for 1 job: 10,96 seconds

- average execution time for 10 job: 3,21 seconds
- average execution time for 100 job: 3,14 seconds

This is explained by the small number of worker nodes used in this experimentation. With few Worker Nodes, a high level of parallelization is not possible and the communication overheads became more relevant. However, the achieved results showed the benefit of such a system and its usefulness for DILIGENT.
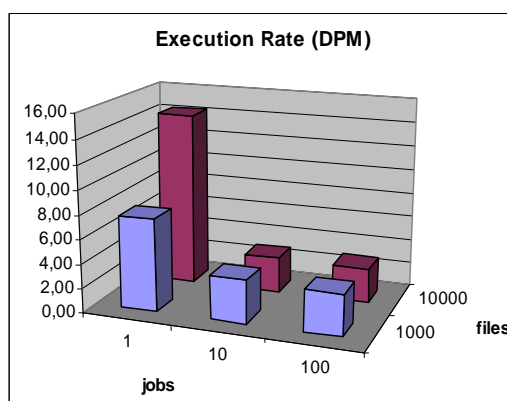

Figure 10: Results for Job Submission

3.1.2.3   gLite 3.0 experimentation

EGEE has made available a new release of its middleware in the end of April. This release, under the name gLite 3.0, is a combination of the two previous software stacks: LCG 2.7 and gLite 1.5. From gLite 3.0 on, there is only one production release. DILIGENT developments and infrastructures were based on gLite 1.4.1. Due to several middleware requirements, DILIGENT decided to move to this new release. To allow a smooth transition, an early experimentation period with the components that are introduced in this release was planned. More detailed information about this activity can be found in the following link:

For a better understanding of what gLite 3.0 includes the following table was prepared:

Service	Component	gLite 1.4	gLite 1.5	gLite 3.0
Access	UI	✔	✔	✔
Secur.	VOMS	✔	✔	✔
Data	DPM	✘	✔	✔
	Fireman	✔	✔	✘
	LFC	✘	✔	✔
	MC	✔	✘	✘
	AMGA MD	✘	✔	✘
	Hydra	✘	✔	✘
	gLite I/O	✔	✔	✘
	FTS	✔	✔	✔
	FTA	✔	✔	✔
	FPS	✔	*	*
Info.	R-GMA	✔	✔	✔
	SD	✔	✔	✔
Job	gLite WMS	✔	✔	✔
	gLite L&B	✔	✔	✔
	gLite CE	✔	✔	✔
	LCG WMS	✘	✘	✔
	LCG L&B	✘	✘	✔
	LCG CE	✘	✘	✔
	Torque	✔	✔	✔
	DGAS	✘	✔	✘
	G-PBox	✘	✔	✘

* FPS is now part of FTS package.

Table 11: Available components in different gLite releases.

AMGA MD, Hydra, DGAS and G-PBox are planned to be included in future gLite releases.

Work Plan

From table 11, DILIGENT has identified a set of tasks (and corresponding components) that should be the focus of the work. A responsible for each task was assigned:

- replacement of Fireman by LFC catalog:
    - CNR - as responsible for the Keeper Service
    - UNIBAS - as responsible for the Content Management Service
- replacement of gLiteI/O by GFAL:
    - CNR - as responsible for the Keeper Service
    - UNIBAS - as responsible for the Content Management Service
- adoption of AMGA:
    - FhG - as responsible for the Metadata Management Service
- interoperability between gLite WMS and LCG WMS
    - CERN
- adoption of G-PBox:
    - CERN (Pedro Andrade) - as DILIGENT representative in the SJ workgroup where this service is being studied
    - UNIBAS (Christoph Langguth) - as DILIGENT representative in the SJ workgroup where this service is being studied
- adoption of BD-II:
    - CERN

It was decided to execute these tasks either by installing the corresponding components in the DILIGENT development infrastructure or by accessing the same components in the EGEE PPS infrastructure. A time plan to perform these tasks was approved:

- 23 Feb: phone conference to introduce gLite 3.0 and distribution of work
- 17 Mar: results sent to the mailing list
- 23 Mar: meeting with Cal Loomis to discuss the outcome of this work
- 27 Mar: TCom presentation about "gLite 3.x adoption in DILIGENT "
- 30 Apr: start planning the upgrade of the DILIGENT infrastructures to gLite 3.0

Results

As planned, the results of the work plan presented above were presented in one TCom conference call. They can be summarized as follows:

General Issues

Several issues regarding the deployment model and the middleware supported were identified:

- The deployment of gLite3.0 can be done in two ways:
    - using the existing xml configuration files and python scripts
    - using the LCG deployment tool called YAIM (also RPM based)
- In gLite 3.0, the middleware support teams will cover the following services:
    - Castor2
    - dCache and DPM

- o FTS
- o LFC
- o GLUE
- o GridICE
- o BDII
- o R-GMA
- o VOMS
- o WMS
- o LB
- o CE

- Fireman and gLiteI/O are not included in gLite 3.0. However they are available as part of gLite 1.5 and have limited support.
- DGAS, Hydra and AMGA are also not included in gLite 3.0. However they are available as part of gLite 1.5 and have limited support. They will be further tested and if proved they can be used in production will be part of future gLite 3.x releases.

LFC

The LFC catalog (already available in previous releases of the LCG middleware stack) replaces the Fireman catalog. The architecture of the two catalogs is quite similar.

LFC provides two interfaces:

- CLI:

 http://grid-deployment.web.cern.ch/grid-deployment/documentation/LFC_DPM/lfc/html
- C API:

 http://grid-deployment.web.cern.ch/grid-deployment/documentation/LFC_DPM/lfc/html

The following table shows the main differences between Fireman and LFC:

	LFC	**Fireman**
WebService Interface	no	yes
Hierarchical Filesystem view of entries	yes	yes
Bulk calls	yes	yes
Oracle backend	yes	yes
MySQL backend	yes	yes
Unix Permissions and POSIX ACLs	yes	yes
VOMS integration	yes	yes

GFAL Library

The GFAL library (also already available in previous releases of the LCG middleware stack) replaces gLiteI/O. Again, the architecture of the two components is quite similar.

GFAL also provides on two interfaces:

- CLI: http://grid-deployment.web.cern.ch/grid-deployment/gis/GFAL/GFALindex.html
- C API

LCG_Util

Is a high-level data management tools to interact with LFC catalog and different types of Storage Elements.

LCG_util provides 3 interfaces:

- CLI: man lcg_util
- C API
- JAVA API: http://grid02.rcub.bg.ac.yu/LFCJavaAPI/files/docs/javadoc/index.html

This tool was used to perform all the data management experimentations. Using the LCG_util Java API it was possible to tests both the LFC client and the GFAL client. The following Java classes where implemented:

- JAVA class to upload local files to a DPM SE and update the LFC catalog
- JAVA class to download files from a DPM SE and update the LFC catalog
- JAVA class to delete files from a DPM SE and update the LFC catalog
- JAVA class to create directories in the LFC catalog
- updated JNI c file to include mkdir call from java API (that was not available)

AMGA

The real functionalities that this metadata catalog provides were still not clear for DILIGENT. Therefore, instead of deploying and experimenting with it, a number of key topics were identified and asked to the AMGA developers. The more relevant questions/answers are now presented:

- Is AMGA fully integrated into gLite and its security/authentication system?
  - Yes. AMGA supports authentication via VOMS including roles using Grid-Proxy certs or MyProxy certificates. Additionally users/groups can be managed within AMGA itself.
- How is the loss of time because of security handshaking solved?
  - AMGA is a stateful service. This means that it can support sessions or even persistent connections. Once a session or connection is established the overhead is virtually 0. Timeouts and quotas are used to prevent individual users from a denial of service attack. Sessions are stored in shared memory or a file-based database.
- To search across collections, are complex queries with joins possible?
  - Yes.
- Will there be support for more complex Schemas, such as XSD/DTD?
  - Anything describable by SQL is supported. A translation from XSD/DTD is not implemented but we would be happy to work with people from Diligent on this issue. It should be technically possible since many DBs storing XML defined data actually use an SQL backend. These include postgresql, MySQL and Oracle which are supported by AMGA as backends. Probably running e.g. Postgres in XML mode would already be half the solution. The devil could be in the detail, though, and some development would be needed I am sure like adding support in the query parser.
- Is service and data replication possible? If yes, is the replication policy scalable and effective?

- o Replication is implemented and available in AMGA 1.1, but there is no security implemented for this feature in that version. Security will be available in 1.2 which is being released within the next couple of days. The implementation's design is supposed to be very scalable but so far only preliminary internal benchmarks were done (which looked very promising). Of course we are interested in any independent tests! Some of these will be forthcoming in the near future from the Gilda team. More important than replication performance is probably the ease of administration and recovery from failures. Both of these are currently heavily worked on and will be shaped according to the feedback we get from the users.

- Is it possible to relate entries in the Metadata Catalog with each other? To create and check Foreign Keys between entries?
    - o Yes, although the creation of the schema may need manual intervention on the backend. Several groups are using these features and it works very well if the schema is not changed dynamically. Full support could be added easily on request.

- Does AGMA support Views?
    - o Yes, on backends which support them.

- Are import tools available?
    - o There is support for dumps and imports of dumps. With the right translation everything can be imported.

WMS

In gLite 3.0, the Job Management components available in LCG 2.7 and in gLite 1.5 are both available. Since DILIGENT relies on the gLite version of WMS a study to identify if the two systems are interoperable was needed. The results are reported in the following picture:

LCG   gLite
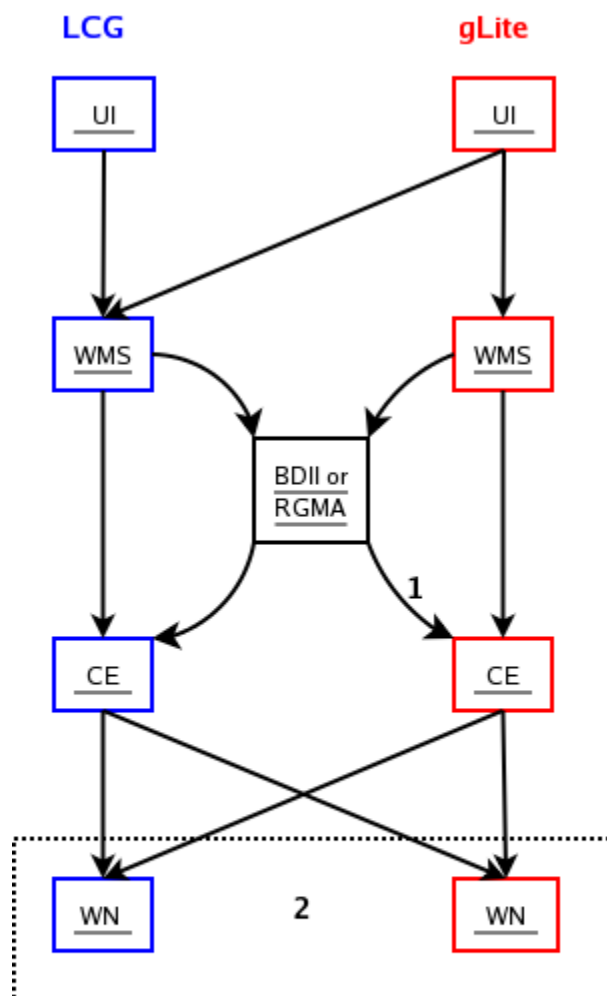
1 - If job is LCG, it will fail. Solution is to force LCG
    jobs to run on LCG CEs (can be done in the JDL).

2 - Submission is possible from LCG and gLite CEs to
    LCG and gLite WNs. In the future, WNs will be seen
    as one. No diference between LCG and gLite WNs.

G-PBox

Stands for Grid-Policy Box and implements policies (rules) on the available resources of a production Grid by:

- introducing policies on a site which are meant to stay local
- creating policies to be transmitted to other sites
- interfacing with decision points (WMS, CE, SE, etc) to merge the information before taking a decision on a resource :
  - WMS: choose the appropriate CE queue
  - SE: decide whether or not to allow the user to execute operations based on things like user-, group- and VO- quotas, membership on VOMS groups, belonging to VOMS roles, etc...

PBoxes are the basic elements of G-PBox, with one installation (at least) for each administrative domain that:

- create and distribute policies (through a policies propagation mechanism)
- receive and evaluate requests

G-PBox provides on 3 interfaces: C API, C++ API and JAVA API.

In the typical G-PBox scenario, a user wants to take some action on a particular resource (e.g. RB, CE, SE). The user acts through a Policy Administration Tool (PAT) to submit its query to the entity protecting a resource which is called a Policy Enforcement Point (PEP). The PEP forms a request (using the XACML request language) based on attributes (subject, action, resource, and other information. The PEP then sends this request to a Policy Decision Point (PDP), which examines the request, contacts the Policy Repository (PR) to retrieves policies (written in the XACML policy language) that are applicable to this request, and determines whether access should be granted according to the XACML rules for evaluating policies. That answer (expressed in the XACML response language) is returned to the PEP, which can allows or denies access to the requester.
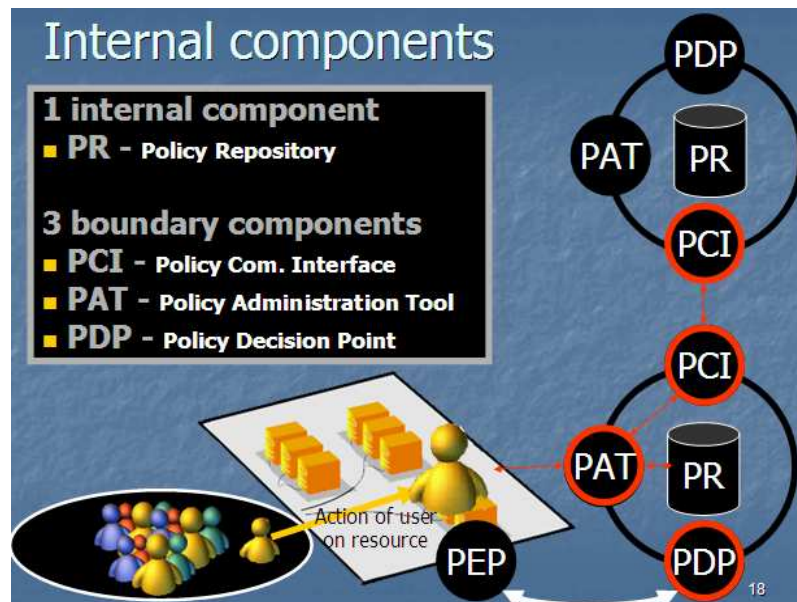
The G-PBox component was mainly evaluated to understand its usefulness to address the DILIGENT requirement for defining job priorities. G-PBox as a policy enforcer can be used to evaluate any type of policies, such as job priorities. The advantage of using G-PBox for job priorities is that instead of having to go through all sites to update the priorities, a single update in G-PBox is enough. This is simpler and clearly less error-prone. The disadvantage is the complexity and the effort needed to set up such system.

BD-II

The BD-II information system (also already available in previous releases of the LCG middleware stack) is the official IS of the EGEE production infrastructure.

This information system relies on the following architecture:

- one local information provider (GRIS) running in each service

- o example: ldap://server1:2135 - mds-vo-name=local,o=grid
- one site BDII or GIIS for each site
  - o example: ldap://server1:2170 - mds-vo-name='sitename',o=grid
- one top-level BDII for each VO
  - o example: ldap://server2:2170 - mds-vo-name=local,o=grid

Any LDAP client can be used to query BD-II. There is also an high-level application – called lcg-infosites – that hides the LDAP query syntax but only queries part of the system.

From the work done with the new gLite 3.0 components the following requirements were identified:

Service	Description	Priority	Deadline
LFC	web service interface for LFC Catalog	critical	
	java API for LFC catalog	critical	
	performance of java wrapper	desirable	
GFAL	java API for service	critical	
	RGMA instead of BDII to query DPM and LFC	desirable	
AGMA	complex schemas	critical	06/06
	relationship between different metadata entries	critical	06/06
	views	desirable	09/06
	OAI	desirable	12/06

3.1.3   Bugs

Bug reporting is an important activity in the interaction between DILIGENT and EGEE. DILIGENT is an early adopter of gLite and therefore the DILIGENT bug reporting activity is an important feedback that is provided to EGEE.

With the beginning of the DILIGENT development phase, the early experimentations with gLite were gradually replaced by the concrete implementation of the DILIGENT services. As a consequence, the number of discovered and reported bugs have increase.

For some of these bugs, the discussion through the EGEE mailing lists was enough. In particular, the following mailing lists were used:

glite-discuss@cern.ch

project-diligent-glite@cern.ch

These mailing lists allowed DILIGENT developers to interact with the gLite developers and testers trying to clarify and give feedback on the problems found. From this discussion, whenever a concrete problem was identified, and it required the intervention of the EGEE development or integration teams, the bug was reported using the official gLite bug tracking tool:

http://savannah.cern.ch/bugs/?group=jra1mdw

The following bugs were identified and submitted to savannah:

1)
https://savannah.cern.ch/bugs/?func=detailitem&item_id=6807
Submitted by: Pedro Andrade <prodrigu>
Submitted on: 2005-02-14 17:04
Category: gLite WMS
Status: Fixed
Open/Closed: Closed
Release:
Summary: DAG job problem

2)
https://savannah.cern.ch/bugs/?func=detailitem&item_id=9205
Submitted by: Stephen Burke <burke> (on behalf of Christoph Langguth)
Submitted on: 2005-06-24 13:00
Category: Data Management
Category: Security
Status: Fixed
Open/Closed: Closed
Release:
Summary: Fireman should not limit the length of the DN

3)
https://savannah.cern.ch/bugs/?func=detailitem&item_id=10313
Submitted by: Pedro Andrade <prodrigu>
Submitted on: 2005-08-17 14:53
Status: Fixed
Open/Closed: Closed
Release: gLite 1.4.1
Summary: VOMS new user - invalid email

4)
https://savannah.cern.ch/bugs/?func=detailitem&item_id=10658
Submitted by: andrea manzi <andreamanzi>
Submitted on: 2005-09-06 15:37
Category: Installation and configuration
Status: Remind
Open/Closed: Open
Release: gLite 1.3
Summary: error on gLite CE 1.3 configuration scripts

5)

https://savannah.cern.ch/bugs/?func=detailitem&item_id=10730
Submitted by: andrea manzi <andreamanzi>
Submitted on: 2005-09-12 11:29
Category: Logging and Bookkeeping
Status: Irreproducible
Open/Closed: Closed
Release: gLite 1.3
Summary: Error on glite-job-submit

6)
https://savannah.cern.ch/bugs/?func=detailitem&item_id=12712
Submitted by: Davide Bernardini <davidebernardini>
Submitted on: 2005-10-28 15:31
Category: Data Management
Status: Ready for Test
Open/Closed: Open
Release: gLite 1.5
Summary: gLite I/O - Dead links inside Fireman Catalog after a failed gLite-put

7)
https://savannah.cern.ch/bugs/?func=detailitem&item_id=13988
Submitted by: andrea manzi <andreamanzi>
Submitted on: 2005-11-30 16:29
Category: Logging and Bookkeeping
Status: Accepted
Open/Closed: Open
Release: gLite 1.4.1
Summary: Failed to start glite-lb-locallogger on gLite 1.4.1

8)
https://savannah.cern.ch/bugs/?func=detailitem&item_id=14445
Submitted by: Henri Avancini <avancini>
Submitted on: 2006-01-10 16:48
Category: Installation and configuration
Status: Ready for Review
Open/Closed: Open
Release: gLite 1.4.1
Summary: CE installation script fails on RH 3 ES (32 bit)

9)
https://savannah.cern.ch/bugs/?func=detailitem&item_id=15022
Submitted by: Alessandro Maraschini <maraska>
Submitted on: 2006-02-16 14:03

Category: gLite WMS

Status: Ready for Test

Open/Closed: Open

Release: gLite 1.4 & 1.5

Summary: wms-ui java getOutput not working

As it can be seen, most of the bugs are either "Close" or "Open - Ready for Test". This means that bugs submitted by DILIGENT were considered effective and actually addressed by EGEE.

3.1.4   DILIGENT Middleware Requirements

Requirements definition

DILIGENT requirements have been defined, discussed and ranked inside the DILIGENT consortium starting from April 2005, when the study of gLite middleware reached an advanced stage. The first steps consisted in getting access to the existing list of requirements formulated by other EGEE user communities and trying to identify and prioritise those shared by DILIGENT.

Later, DILIGENT partners experimented with the Grid middleware in a real user scenario and started the detailed design of the services to be developed. Immediately after, a coordinated experimentation of gLite new functionalities was carried on by the technical partners. These experiences gave the opportunity to investigate on specific functionality of the gLite software and set the foundation for the definition of more advanced requirements.

Requirements submission

During 2005 EGEE requirements were collected, prioritized and managed into a central database, managed by the EGEE PTF. More information about these requirements can be found at:

<p style="text-align:center">https://svn.lal.in2p3.fr/EGEE/PTF/web/requirements.html</p>

The EGEE Project Technical Forum (PTF) involved representatives from all activities and was intended to be a forum where a consensus on technical issues could be forged. One of the tasks of the PTF was to track the EGEE project requirements. Requirements were submitted by the user communities to the requirements database through a designated administrator. PTF members had then to analyse, accept/refuse and rank the requirements.

DILIGENT appointed a requirements manager for the project who was authorized to access to the EGEE requirements database. The DILIGENT requirements manager could express the project's interest in the resolution of a selected number of existing requirements (specifying the priority).

Collaboration with EGEE TCG and workgroups

During the second half of year 2005, the responsibility of dealing with user requirements passed from the PTF to the EGEE Technical Coordination Group (TCG). More information on the EGEE TCG at:

<p style="text-align:center">http://egee-intranet.web.cern.ch/egee-intranet/NA1/TCG/tcg.htm</p>

The TCG brings together the technical activities within the project in order to ensure the oversight and coordination of the technical direction of the project (making sure that SA1, SA2, SA3, NA4 and JRA1 have a common view), and to ensure that the technical work progresses according to plan. The TCG has responsibilities in three areas: service APIs, requirements, and technical relationships with outside organisations. By design the group contains representatives from all EGEE activities. The group receives and clarifies input requirements from the activities, the steering boards and from short-lived workgroups that define goals in certain areas (e.g. data management, security, monitoring, etc.). The group then prepares, together with the activity leaders, the agreed work plans for short periods and ensures that the plans are realistic, compatible and match what is needed. One of the roles of the group is to document and publish the discussions and decisions. The requirements of different user communities and the expectations of different resource centres, as service providers, are coordinated by the TCG. Such requirements should be effectively incorporated in the deployment of the Grid infrastructure and associated Grid middleware in a cyclical manner throughout the project's lifetime.

DILIGENT has established a close collaboration with the EGEE TCG though its TCG representative (NA4 activity leader Cal Loomis). DILIGENT members are in contact and regularly meet with the DILIGENT TCG representative. The general objective of these meetings is to raise issues related to the adoption of gLite in DILIGENT and understand about EGEE future plans for development and support.

The identification of the DILIGENT requirements follows the analysis of the typical DL scenario that requires the repetition of the same operations for each input files belonging to a given collection. Each single operation takes few minutes (or less) but it must be repeated for hundreds of thousands times (even millions). These operations are realised as executables and normally must be organised in a workflow to deliver a more complex functionality designed to make accessible a new collection of objects through the DILIGENT application. Each object belonging to a collection must be stored in a SE, must be automatically described with metadata and features automatically extracted using the appropriate executables; the metadata and the features must be analyzed and indices must be created to improve the accessibility of these objects and these operations are done using other appropriate executables. The created indices are managed by DILIGENT services but are also persisted in SEs in order to allow the relocation of the services. Moreover, other operations can be required by the users: encryption of the objects and/or watermarking, trans-coding and compression, etc. All these operations are managed by the DILIGENT application that for each collection will prepare an appropriate workflow that takes into account the requirements of the user, the type of the objects, and the available executables with respect to the user rights. Any time a new DL is created many collections could be managed, and many DLs could be created at the same time.

In order to support this general framework, it should be possible to:
- query for the maximum number of CPUs concurrently available,
- use parametric jobs,
- support service certificate,
- specify a job specific priority,
- specify a priority for a user or for a service,
- ask for on-disk encryption of data,
- dynamically manage VO creation,
- dynamically support user/service affiliation,
- support multiple VOs affiliation.

The detailed motivations and rationales of these requirements are contained in the following document:

https://uimon.cern.ch/twiki/pub/DILIGENT/DiligentRequirements/DiligentVision.doc

The relationships of these requirements with the ones expressed by the DILIGENT project through the PTF database are accessible at:

https://uimon.cern.ch/twiki/bin/view/DILIGENT/DiligentRequirements

These requirements are presented and discussed with TCG representatives, see event #45. DILIGENT members are rooted in joining a number of EGEE workgroups active in specific areas of interest for the project. In particular DILIGENT has representatives in the following workgroups:

- Short Deadline Jobs
- Medical Data Management
- Job Priorities
- Middleware Security Workgroup
- Joint Security Policy Group

DILIGENT members participated in all phone conferences and meetings organised so far by these working groups, see event #42, #43, #48, #49, and #52.

3.2   Interactions related to Networking

DILIGENT and EGEE had a considerable number of contacts through which the synergies among the two projects were successfully improved. Thanks to the exploitation of the instruments provided by the EGEE networking activity (web portal, training materials and events), DILIGENT members improved their knowledge and expertise about Grid technology. In this second phase of the project, they changed their roles moving from listeners to contributors and reporting the results of their own experimentation activities in conferences and workshop organised by EGEE. Nevertheless, DILIGENT partners continue to attend EGEE projects events (e.g. #44, #47) to be informed about future plans, current priorities, adopted best practises, difficulties, etc.

DILIGENT is also establishing relations and collaboration agreements with other EGEE related projects, among them ETICS and ArcheoGrid.

gLite Installation and Configuration Training (#29), Geneva, CERN, 16-17 June 2005

Topic/Purpose: gLite installation and configuration tutorial.

Participants: Pasquale Pagano (CNR-ISTI, DILIGENT), Manuele Simi (CNR-ISTI, DILIGENT), Alessandro Tomei (ENG, DILIGENT), Konstas Tsakalozos (UoA, DILIGENT), EGEE trainers.

Agenda:
Day 1 - Introduction & Hands-on session
Day 2 - Hands-on session

Results: The first tutorial about the installation and configuration of the gLite middleware allowed DILIGENT members to improve their knowledge about the middleware. Exploiting this knowledge, the first DILIGENT infrastructure was improved and tested with more sophisticated testing activities. The complete result of these testing activities is reported at the following link:

https://uimon.cern.ch/twiki/bin/view/DILIGENT/GLiteDiligentTC

4th EGEE Conference (#37, #38), Pisa, October 2005

Topic/Purpose: DILIGENT participation at the 4th EGEE project conference: Global and Persistent e-Infrastructure for Scientific Knowledge in the 21st Century. The theme of the conference was creating and managing knowledge in the scientific community and building solutions which can be exported to industry and commerce. Workshops brought together delegates from many different countries, providing them with the opportunity to share knowledge, experiences and discuss the future. The conference also gave members of the project the opportunity to evaluate their achievements, prepare for the final project review and plan for the next phase of the project.

Participants: The complete list of participants can be accessed at the EGEE conference web site available at: http://public.eu-egee.org/conferences/4th/

Agenda: The agenda of the 4th EGEE was pretty reach. The DILIGENT project was present with a number of technical people covering the different areas. DILIGENT members presented DILIGENT at the plenary session and during a technical parallel session. In particular:

- the project Scientific Coordinator Donatella Castelli presented "DILIGENT: Objectives & Current Status" and participated to a panel discussion on "Grid and EGEE: The Digital Library perspective"
- "DILIGENT Technical overview" was presented by DILIGENT Technical Support Manager Pasquale Pagano and Pedro Andrade
- DILIGENT and EGEE members participated to an open technical discussion on DILIGENT requirements

Results: There was a plenary talk by Donatella Castelli where she spoke of the current activities of DILIGENT, gLite usage of DILIGENT and future plans. The results of the gLite experimentation were presented in an NA5 (Collaborations) session. DILIGENT also organised a JRA1 & JRA3 technical meeting during the conference to discuss these results in detail and to highlight what is needed in the future with regards to the middleware functionality.

EGEE NA3 internal review (#40), NeSC, Edinburgh, UK, November 2005

Topic/Purpose: Peer review of the EGEE training activities performed in the previous year

Participants: Malcolm Atkinson (NeSC, EGEE), David Fergusson (NeSC, EGEE), Bon Low (NeSC, EGEE), Andrea Manieri (ENG., DILIGENT).

Agenda: The meeting presented the courses given, the shared training material, the work performed during the Summer Schools, the T-Infrastructure and e-Learning facilities, and the result of the requirements analysis. The detailed agenda can be found at the following address:

http://agenda.cern.ch/fullAgenda.php?ida=a056662

Results: The reviewers have discussed about the missing or useless actions; they pointed out the needs to improve the collaboration between projects, and made suggestions about the EGEE eLearning initiative and EGEE II.

Meeting with EGEE NA4 (#41), Geneva, 30 November 2005

Topic/Purpose: Meeting with NA4 leaders to discuss the future collaboration between DILIGENT and EGEE NA4

Participants: Pedro Andrade (CEN), Florida Estrella (CERN), Vincent Breton (EGEE), Massimo Lamanna (EGEE), Roberta Barbera (EGEE, on the phone), Cal Loomis (EGEE, on the phone)

Agenda:

- Procedure on how DILIGENT collaborates with NA4
- Procedure on how to express requirements
- DILIGENT participation in NA4 events

Results:

A Memorandum of Understanding (MoU) must be signed between DILIGENT & NA4. Currently NA4 only has MoUs with user communities that want to deploying applications on the EGEE infrastructure. DILIGENT is the 1st external project collaborating project with NA4, so DILIGENT MoU might serve as a model or template for the other new EU projects. The MoU should include all activities of EGEE that DILIGENT will make use of (e.g. JRA1, SA1, NA2, NA3, NA4, etc), what DILIGENT wants to do with the EGEE middleware, what are the milestones & plans, the requirements that need to be addressed and also the manpower implications on EGEE. The MoU may be submitted to the EU, perhaps as an appendix for the WP3.1 deliverable.

Cal Loomis is from now on the NA4 contact with respect to requirements. The requirements database (PTF database) is still available and although it has not been active for a couple months it is still accessible for update. The TCG is taking it in consideration and is responsible for organizing, prioritizing, and addressing the requirements into working groups.

Roberta Barbera presented and invited DILIGENT to the NA4 Generic Application meeting organised in Catania, Italy in the beginning of January 2005.

EGEE NA4 Generic Applications Meeting (#46), Catania, 9-11 January 2006

Topic/Purpose: The meeting was devoted to the presentations of the different Generic Applications.

Participants: Andrea Manzi (CNR, DILIGENT). The complete list of participants is available at:

http://egee-na4.ct.infn.it/NA4-meeting06/participants.php

Agenda: In each presentation, the speaker included the description of the activities performed exploiting the EGEE middleware, the use of the EGEE production infrastructure, highlighting main issues and problems. The complete agenda of the meeting can be accessed at the meeting Web site available at:

http://agenda.cern.ch/fullAgenda.php?ida=a058080

Results: DILIGENT was presented to the generic applications audience. After a brief introduction of the project goal and expectations, the result obtained exploiting the gLite middleware was presented, the problems encountered with this middleware was deeply explained, and the main requirements were illustrated.

EGEE User Forum (#51), Geneva, 1-3 March 2006

Topic/Purpose: The EGEE User Forum provided an important opportunity for innovative applications to establish contacts with EGEE and with other user communities, to learn about the latest advances, and to discuss the future evolution in the Grid middleware.

Participants: Andrea Manzi (CNR, DILIGENT), Davide Bernardini (CNR, DILIGENT), Christoph Langguth (UNIBAS, DILIGENT), Laura Voicu (UNIBAS, DILIGENT), Pedro Andrade (CERN, DILIGENT). The complete list of participants is available at:

http://indico.cern.ch/confRegistrantsDisplay.py/list?confId=286

Agenda: The complete agenda of the meeting can be accessed at the meeting Web site available at:

http://indico.cern.ch/conferenceTimeTable.py?confId=286&showDate=all&showSession=all&detailLevel=contribution&viewMode=parallel

Results: DILIGENT was represented in all main parallel sessions. This allowed DILIGENT project being aware of the most important activities currently running in the Grid environment. Moreover, DILIGENT was presented by the CNR in a comparison with the currently feasible exploitation of the Grid technology in a presentation entitled "DILIGENT and OpenDLib: long and short term exploitation of a gLite Grid Infrastructure"

Meeting with EGEE NA4 (#54), Geneva, 12 April 2006

Topic/Purpose: Meeting with NA4 leader Cal Loomis to discuss the work done by TCG, the work done by DILIGENT in the TCG workgroups, DILIGENT requirements for gLite3.0, the new organisation of NA4 and the need for DILIGENT MoU in the context of EGEE-II

Participants: Pedro Andrade (CERN, DILIGENT), Roberta Faggian Marque (CERN, DILIGENT), Charles Loomis (NA4 activity leader and DILIGENT representative in TCG)

Agenda:

- Presentation of DILIGENT requirements
- Outcome of last TCG meetings
- EGEE development and support plans
- Need for a DILIGENT-EGEE MoU

Requirements presented:

Besides the requirements being analyzed by the TCG workgroups, with the introduction of new components in gLite 3.0 also new requirements were identified by DILIGENT. These requirements were explained to C. Loomis:

- web service interface for LFC
- java API for LFC
- java API for the LCG utilities
- possibility to use R-GMA (instead of BDII) to publish or retrieve information about DPM and LFC from the LCG utilities

Results:

- C. Loomis took note of DILIGENT major requirements and discussed possible solutions for the usage of R-GMA to retrieve information about DPM and LFC from the LCG utilities. Since for know this task is impossible, C. Loomis proposed to have both information system running, continue to use R-GMA as the default IS and automatically duplicate the information in BDII using existing gLite tools.
- C. Loomis informed that there is not a clear procedure for which the project could be informed about the timeline for implementation of TCG resolutions since this is in the hands of the single activities.
- C. Loomis clarified on EGEE plans for the future: no new functionality will be developed since focus will be given to stability of the current system. Minor features could be implemented if required by the project.
- C. Loomis informed on the existence of a procedure by which new external components could be integrated into the gLite software stack. This means that if DILIGENT identifies external components which could provide important functionality, these components could be submitted to this integration process.
- C. Loomis clarified on the level of support that will be provided to the different gLite components.
- C. Loomis explained about the new EGEE procedure for interaction with user VOs. A VO should register though a Web interface to become an EGEE registered VO and get access to training and test infrastructure. A VO could ask for specific support and access to production infrastructure. In this case the VO request will be submitted to approval and the VO will become a selected VO. Specific collaboration and support condition could be agreed in a Memorandum of Understanding (MoU).
- It was agreed that a MoU could be useful for DILIGENT to agree on the following aspects: integration of DILIGENT with EGEE production infrastructure, needed support from EGEE operation team, resources needed and provided by DILIGENT, possible EGEE training to be provided on specific areas and joint technical meetings to discuss about DILIGENT specific design choices which involve exploitation of gLite functionalities, gLite software short and long term support needs from DILIGENT.

3.3 Interactions related to Operations

The interactions related to operations between DILIGENT and EGEE have made important progresses. If in the first period DILIGENT was only focused in the setting up its own infrastructure, in the second phase besides the regular update of DILIGENT infrastructures, DILIGENT got access and started to use the EGEE Pre-Production Service Infrastructure. In addition, the discussion between EGEE and DILIGENT on how to connect their two infrastructures has also started.

3.3.1 DILIGENT Infrastructures

DILIGENT currently maintains two independent infrastructures: the development infrastructure and the testing infrastructure – this means, infrastructures only based on DILIGENT resources. Although they are not connected to any EGEE infrastructure the collaboration with EGEE is extremely important to allow the proper set up and maintenance of the infrastructures.

During this reporting period, DILIGENT continued to work on the set up, upgrade and maintenance of these infrastructures. They were initially installed with gLite release 1.1 and after upgraded two times. A third upgrade is being planned:

Installation of gLite 1.1
- When: June 2005
- Infrastructures: DILIGENT development and DILIGENT testing
- Description: CERN and CNR coordinated this activity. This corresponds to the first coordinated installation of two DILIGENT infrastructures in some DILIGENT sites. Not all services were installed. Only around 50% of the services were made available due to some bugs found in this release.
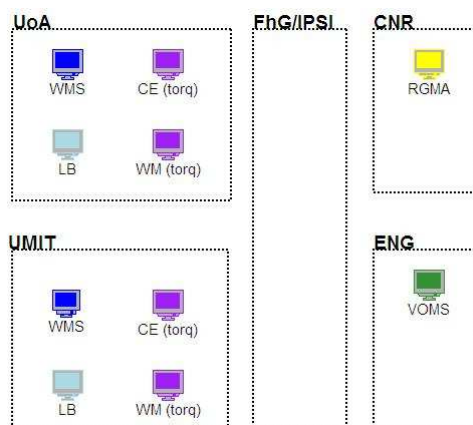

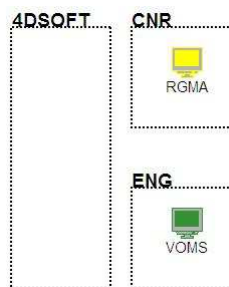*Figure 11: DILIGENT development infrastructure running gLite 1.1*

Figure 12: DILIGENT testing infrastructure running gLite 1.1

Upgrade from gLite 1.1 to gLite 1.3

- When: August 2005
- Infrastructures: DILIGENT development and DILIGENT testing
- Description: CERN and CNR coordinated this activity. After the upgrade to gLite 1.3, for the first time, all gLite components were made available in both infrastructures.
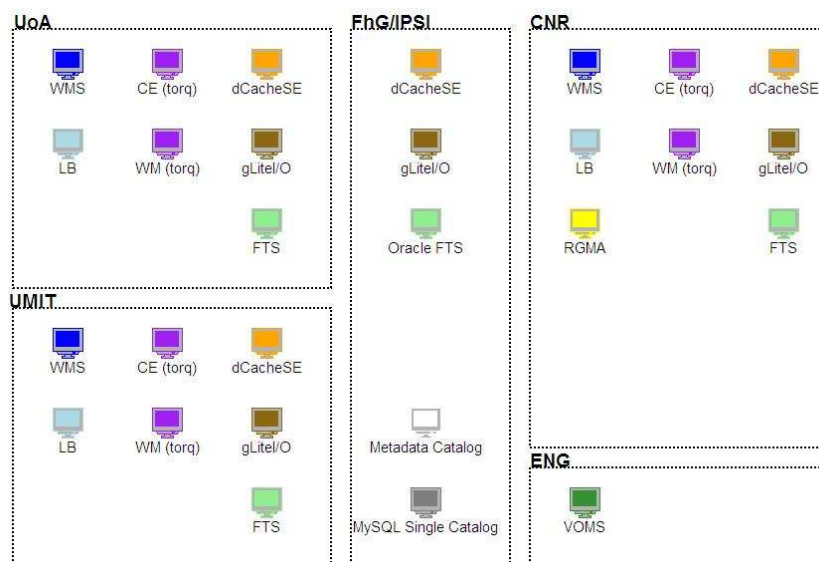

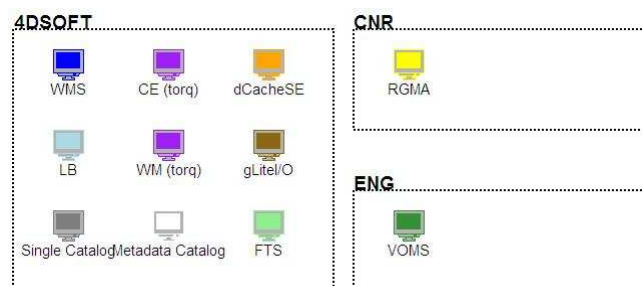*Figure 13: DILIGENT development infrastructure running gLite 1.3*

Figure 14: DILIGENT testing infrastructure running gLite 1.3

Upgrade from gLite 1.3 to gLite 1.4.1

- When: November 2005
- Infrastructures: DILIGENT development and DILIGENT testing
- Description: CERN and CNR coordinated this activity. After an initial upgrade planning phase carried out by CERN and CNR, an upgrade plan was proposed to all DILIGENT technical partners. This plan was accepted and two conference calls were set up to follow the upgrade progresses. The upgrade process took approximately two weeks.
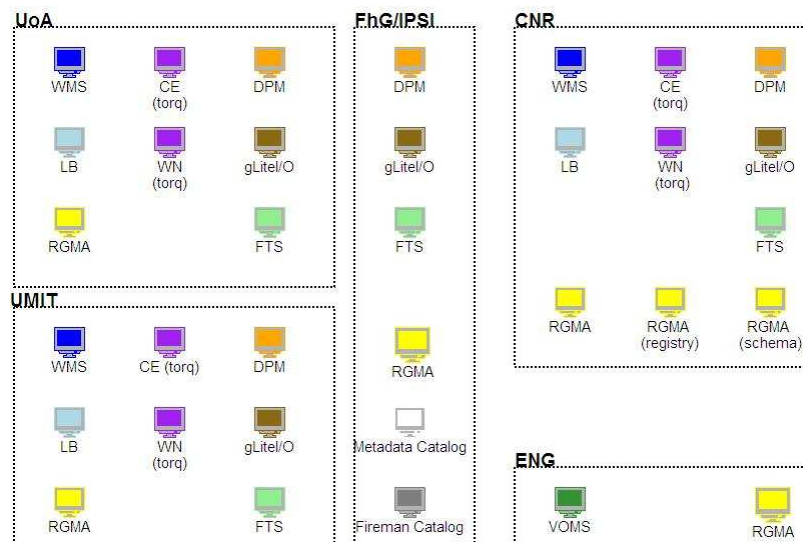

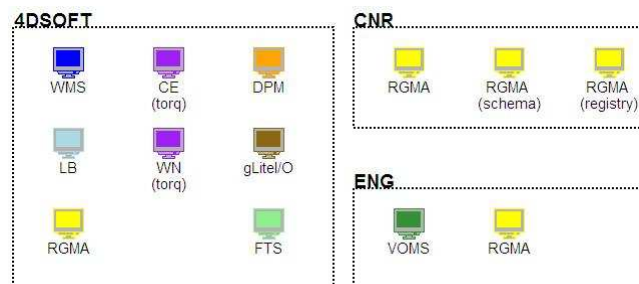*Figure 15: DILIGENT development infrastructure running gLite 1.4.1*

Figure 16: DILIGENT testing infrastructure running gLite 1.4.1

Upgrade from gLite 1.4 to gLite 3.0 (still under definition)

- When: May/June 2005
- Infrastructures: DILIGENT development and DILIGENT testing
- Description: CERN and CNR coordinated this activity. It was decided to upgrade directly from gLite 1.4 to gLite 3.0 and skip the gLite release 1.5 due to a number of reasons:
  - several bugs were found in gLite 1.5 (although QFs are available)
  - gLite 1.5 was released without the usual certification process
  - most DILIGENT effort is now being allocated to the development of the first version of the DILIGENT services

Parallel to install and upgrade work, monitoring of the infrastructures was carried out and an automatic process to check the status of the services running on the infrastructures was developed. The output of this application can be found here:

- http://diligent.web.cern.ch/diligent/gss/dev.html
- http://diligent.web.cern.ch/diligent/gss/test.html
- http://diligent.web.cern.ch/diligent/gss/chart.html

3.3.2 DILIGENT-EGEE Pre-Production Infrastructure

To allow DILIGENT users to access and use the EGEE PPS infrastructure a number of interactions between DILIGENT and EGEE was needed:

- September 2005: Meeting with EGEE SA1 PPS responsible Nicholas Thackray about DILIGENT access to PPS infrastructure (#34)
- September 2005: PPS infrastructure access is granted (#35)
- October 2005: "diligent" VO created and populated with DILIGENT members
- October 2005: Discussion with PPS administrators to coordinate the addition of the "diligent" VO in some PPS sites
- March 2006: PPS infrastructure updated to gLite 3.0. Discussion with PPS administrators to revoke "diligent" VO access to PPS sites

The access to an external infrastructure and in particular to the EGEE PPS is of great importance to the DILIGENT project. This is justified by two major reasons:

- PPS gives to DILIGENT the opportunity to test its developments in an external infrastructure, where DILIGENT has no administrative control. It provides a way to understand if there is any dependency with the DILIGENT infrastructure and to compare the results.
- In the case of possible problems with its own "no-production-quality" infrastructure, DILIGENT can always count on the availability of an external infrastructure to pursue its tests.

Another aspect that promoted the collaboration between the two projects was the creation of a new VO in the PPS for the DILIGENT community. This process required several meetings since, the procedure for adding a new non-EGEE VO was not clear at that time. DILIGENT was actually one of the first "external" VOs in PPS. DILIGENT was therefore a useful for EGEE to understand this process.

Currently, DILIGENT members are using the PPS infrastructure using the "diligent" VO and of the 20 sites that compose the infrastructure DILIGENT has access to 6 of them:

- CERN - https://goc.grid-support.ac.uk/gridsite/gocdb2/index.php?siteSelect=235
- CNAF - https://goc.grid-support.ac.uk/gridsite/gocdb2/index.php?siteSelect=202
- LIP - https://goc.grid-support.ac.uk/gridsite/gocdb2/index.php?siteSelect=201
- PIC - https://goc.grid-support.ac.uk/gridsite/gocdb2/index.php?siteSelect=178
- UPATRAS - https://goc.grid-support.ac.uk/gridsite/gocdb2/index.php?siteSelect=167
- SWITCH - https://goc.grid-support.ac.uk/gridsite/gocdb2/index.php?siteSelect=308

As a consequence of the access to PPS, and besides the usage of the infrastructure itself, other interactions between DILIGENT and EGEE occurred:

- DILIGENT members got access to the PPS Wiki site (private section)
  - https://pps-private-wiki.egee.cesga.es
- DILIGENT members subscribed to the PPS mailing lists
  - project-eu-egee-pre-production-service@cern.ch
  - project-eu-egee-pre-production-users@cern.ch

3.3.3 DILIGENT-EGEE Production Infrastructure

To increase the DILIGENT computing and storage power, DILIGENT plans to connect its infrastructure to the EGEE infrastructure.

With the implementation of the DILIGENT services ongoing the planning of the future DILIGENT production infrastructure and how it can join the EGEE production infrastructure has started.

Although the two infrastructures run the same Grid middleware it is clear that connecting the infrastructures is not a straightforward task. To start identifying the possible issues and the way to process a first meeting between DILIGENT and EGEE took place.

Meeting with EGEE SA3 leader (#57), Geneva, 18th May 2006

Topic/Purpose: Discuss the requirements for connecting the DILIGENT infrastructures with the EGEE production (and pre-production) infrastructures.

Participants: Pedro Andrade, Roberta Faggian Marque, Markus Schulz (EGEE SA3 activity leader)

Results:
- Interoperability:
  - Study of the differences between the DILIGENT extended GLUE schema and EGEE GLUE schema is the first and key step start ASAP. DILIGENT should use GLUE schema 1.2.
  - The DILIGENT Information System (DIS) is R-GMA based while EGEE IS is based on BD-II. R-GMA is used in EGEE just for monitoring.
    - to allow the DIS to work efficiently all information in BDII must be "copied" to RGMA,
    - to allow the usage of tools like LCG_UTILS info about some gLite resources (CE, SE, etc) should also be available in BDII
- General security issues:
  - DILIGENT has to prove it works under a controlled environment.
  - End users who originated the Grid jobs/operation have always to be tracked.
  - End users should sign an "responsibility contract" with DILIGENT.
- DILIGENT services:
  - DILIGENT needs (at least) to deploy its DILIGENT Hosting Node (DHN) in the EGEE sites.
  - DILIGENT needs to explain how it's planning to do the load balancing of its services as well as its bootstrapping process .

- Steps to join EGEE production:
  - simple interoperability tests between DILIGENT and EGEE GLUE schemas,
  - pre-production tests between DILIGENT PPS sites and EGEE PPS sites,
  - DILIGENT joining EGEE production.

Meeting with EGEE SA3 (#58), Geneva, 29ʰ May 2006

Topic/Purpose: Discuss the interoperability issues of connecting the DILIGENT infrastructures with the EGEE production (and pre-production) infrastructures

Participants: Pedro Andrade, Laurence Field (EGEE SA3)

Results:
- Interoperability:
  - DILIGENT is using an extend version of GLUE schema 1.2
  - DILIGENT extended GLUE schema doesn't not interfere with the EGEE GLUE, so no additional translators are needed
  - the DILIGENT Information System (DIS) is R-GMA based while EGEE IS is based on BD-II. R-GMA is used in EGEE just for monitoring.
- Deployment
  - Since the two infrastructures have interoperable information systems, the discussion should now focus in how to organise the deployment of the DILIGENT PPS sites.

3.4 Feedback to EGEE and DL community

The feedback activity resulting from an in depth understanding and usage of the EGEE technologies have seen important progresses. While the first period was devoted to anticipate as much as possible the understanding of the different aspects of the gLite middleware, the second period is characterised by the concrete exploitation and use of this technology. For instance, DILIGENT got access and started to use the EGEE Pre-Production Service Infrastructure as well as the detailed design and implementation of the DILIGENT services started. The experiences gained enabled to produce comments and reactions that are presented in the remaining of this section.

3.4.1 Mechanisms

The mechanisms put in place in the previous phase of the project, i.e. EGEE requirements database, Mailing lists, Web site, Technical meetings, Documentation, and Training, have been complemented by:

- **"Short deadline jobs", "Middleware Security" and "Joint Usage Policy" Working groups**: these have provided a new instrument for extensive, face-to-face, discussions with gLite designers and developers on key topics of interest. These discussions have also stimulated further joint investigation which will be carried out in the next months.

Information Society
Technologies

Diligent
A **DI**gital **L**ibrary **I**nfrastructure
on **G**rid **EN**abled **T**echnology

- **Conferences and thematic events organised by the Grid and DL research communities:** the participation to these events is another important mechanism used in this period for proving feedback to the two communities. It is expected that this form of communication will be largely increased in the near future as more concrete results are achieved.

3.4.2   Feedback to EGEE

Detailed DILIGENT requirement on gLite middleware have been formulated and reported in this document on 3.1.4. This section illustrates a number of limitations of the current Grid technology with respect to the DL application requirements. These limitations have been identified during this second period of the project. The role of DILIGENT is in fact to clearly identify and fill in this gap offering a Digital Library infrastructure which better responds to users' expectations.

Replication: Replication on the Grid is a substantial requirement for reliability, availability, and performance of Grid-enabled DL applications. In a Grid environment, the number of replicas has to be determined dynamically, depending on the access to data and the availability of the hosts data resides on. Although data replication is part of the Grid data management concepts, there is only limited support in current Grid environments. gLite provides basic mechanisms to replicate files in multiple locations that must be enhanced to be exploitable in the DL context where the virtual notions of collection and aggregate types have also to be taken into account. Therefore, additional developments are needed in order to have sophisticated DILIGENT storage management on the Grid that dynamically replicates data among different Grid nodes and that provides a high degree of availability, reliability and performance. In particular, existing cluster replication protocols must be made "Grid-aware" to cope with heterogeneity, site autonomy and network traffic costs on the Grid. For instance, acceptable answering times of Content Management operations can only be achieved if data is physically replicated on Grid nodes that can be efficiently accessed from the invoking service. Any replication component must, thus, include a high degree of adaptability to the current service workload.

Freshness: In DILIGENT, derived data is generated and stored in multiple cases (e.g., in indices, extracted features, metadata). It is of high importance that this derived data is consistent and 'fresh' (i.e., changes in the original data have to immediately trigger the update of derived data). Current approaches however only consider the update of derived data at some fixed time intervals (e.g., once a day). Adding support for the active propagation of derived data and thus increasing the quality of this data would lead to a significant gain for DILIGENT. Search, for instance, will be more accurate when being based on up-to-date metadata since it will be able to also find newly generated data and documents in the system. DILIGENT needs to increase the freshness of data for DL applications by considering an active propagation of changes. For active propagation of changes or newly inserted documents, processes need to be defined and automatically executed.

Responsiveness of the system: the end-users interacting with a Digital Library will have some expectation with respect to the responsiveness of the system; i.e. if a user searches for data or retrieves some data which is generated on-the-fly, the user is actively waiting for the output, so it is important to be able to deliver the output in a timely manner. Such interactivity requirements usually don't apply to the "traditional" domains of Grid computing,

like the processing of massive amounts of high-energy physics data. Therefore the latency of the Grid middleware, for example the time elapsed before a Grid job is started, or the time taken to deliver the results of a job, both ranging in the order of minutes, is an aspect that has been somewhat neglected.

The "Short Deadline Jobs" working group (SDJ WG) is composed by user communities which want to take advantage of the Grid for jobs that are usually not long-running and for which, as a consequence, the latency of the middleware is considered a rather serious restriction. DILIGENT is actively involved in this working group, which strives to develop strategies to reduce the middleware penalty. Possible solutions which are being investigated by the SDJ WG include fine-tuning of the gLite configuration, the deployment of special computing element nodes, and the usage of so-called pilot jobs (in a nutshell, these are normal jobs that, when executed, provide another layer for directly accessing the corresponding worker node and starting jobs with a dramatically reduced latency).

3.4.3   Feedback to the DL Community

The effects of exploiting the Grid technology for supporting DL applications are progressively clarified as the design and experimentation phases advance. The writing of this deliverable report offers the opportunity to formulate the outcome of this experience in a way that could provide interesting feedback also to the DL community. Below some considerations are reported.

Information Retrieval in a Distributed Uncontrolled Environment: DILIGENT attempts to process information on top of a distributed, partially uncontrolled infrastructure whose heterogeneous elements are bound by loose rules. Trying to bring together the power of information retrieval and structured content processing on top of this environment shows that the various typical approaches applied in Digital Library Management Systems and Database Management Systems are not enough to provide a stable and efficient system. Extensive planning of processing operations has to be performed and the activities that come up as a result of a request have to be continuously monitored so that action is taken as soon as failures or inefficiencies are predicted or encountered. Borrowing ideas from the field of Query Optimisation, and appropriately enrich them with dynamic features, seems to be a potential solution to the challenges raised.

Distributed Information Retrieval: Interactions between the DLs and the research field commonly defined by the term Distributed Information Retrieval (DIR) (roughly, probabilistic retrieval of distributed and largely unstructured media content) have been somewhat limited so far. While content-based retrieval engines are a typical offering of classic, centralised DLs – normally in the form of ad-hoc configurations of off-the-shelf tools based on variations of vector-based models (already not exactly the state of the art even for centralised IR) – they are way less commonly found in the modern, fully-distributed projects which characterise so much of the current developments in the DL field worldwide. Distributed IR engines have been the object of research for more than ten years now, mostly in the form of meta-engines/search brokers, and yet there is no well known tool which is under consideration in the DL projects previously mentioned. This is in stark contrast with the case of structured, metadata-based retrieval, in particular the impact that the OAI initiative has instead had on the shaping itself of those projects. There might be many reasons for this unbalance, including the relative complexity/cost of Distributed IR solutions, the potential for sub-optimal network performance of the meta-engine approach, and the lack of proposals which are infrastructural in nature such as open frameworks and

protocols (as opposed to ad-hoc, experimental prototypes). Potentially, the DILIGENT effort may play a role in re-establishing a balance, for the cooperative-nature upon which the underlying GRID is based may absorb many of the complexity/costs which instead discourage more dedicated initiatives. A DILIGENT DL, in particular, makes available state-of-the-art DIR services at no further cost of adoption than those required to participate to the DILIGENT infrastructure. Furthermore, the approach followed in designing and implementing those services is extremely infrastructural in nature, in that the DILIGENT services are designed for generality and ease of extension. Finally, DILIGENT is striving to minimise the impact of network latencies in the way its services operate and interact with other services, in a systematic way that has not been previously approached in the past.

A DL system reference architecture: The vision and the lessons learned during the first part of the project have also strongly influenced the definition of a Reference Architecture for DL Systems introduced by a number of DILIGENT participants as part of the activities of the DELOS Network of Excellence on Digital Libraries. This architecture, depicted in Figure 17 consists of three logical layers: (i) Application Framework, (ii) Enabling Components and (iii) DL Application Components. The DILIGENT infrastructure has a concrete architecture that is compliant with this reference one. In particular, the layers of the Reference Architecture are implemented, respectively, by: (i) gLite and WSRF; (ii) a subset of the DILIGENT Collective layer services and the Process Management ones; and (iii) the services of the three DILIGENT functionality areas, i.e. Content & Metadata Management, Index & Search Management, and User-Community specific applications.
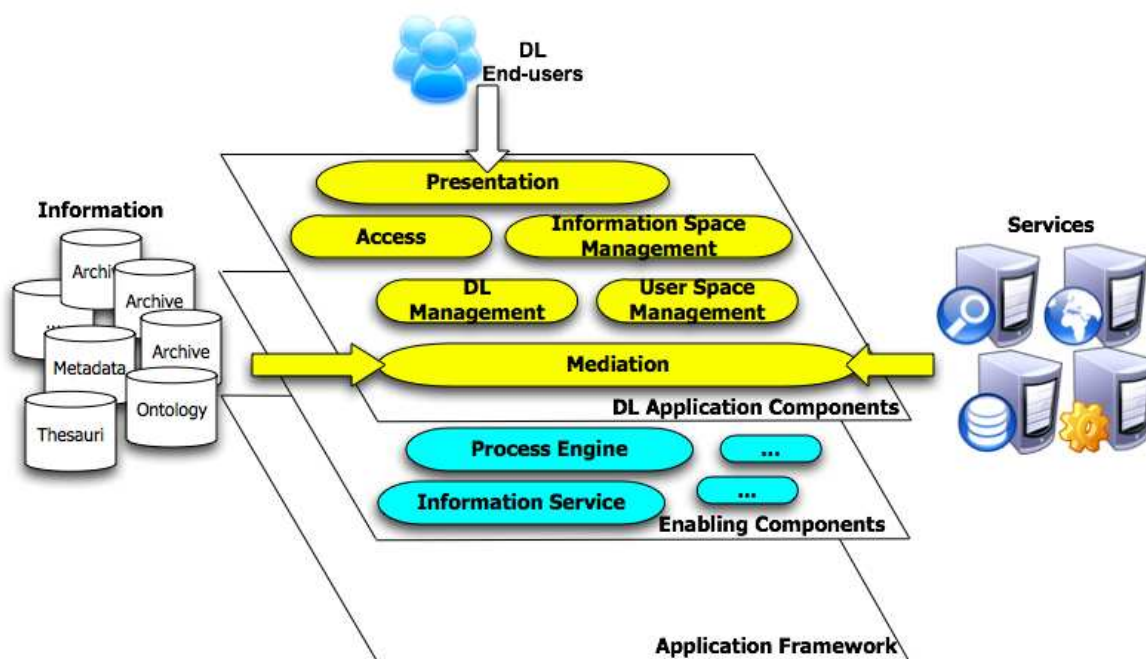

Figure 17: A DL System Reference Model Architecture

The DILIGENT infrastructure is not a DL system but a creator and manager of multiple DL systems sharing the same resources. Therefore, it also implements additional "meta" functionality. This functionality is realized thought the co-operation of the VDL, Keeper and Dynamic VO Support services.

3.4.4 Dependencies between DILIGENT and EGEE

DILIGENT adopts the EGEE middleware as its Grid middleware solution, thus both DILIGENT services and gLite services contribute to compose the DILIGENT infrastructure. Below the main dependencies between DILIGENT and EGEE from two points of view, i.e. software dependencies and infrastructure dependencies, are reported.

Software: The DILIGENT services have been designed to include or wrap parts of the gLite software. In particular the following dependencies exist:

- The DILIGENT Dynamic VO Support Service relies on the gLite VOMS in order to manage the virtual organisation management mechanisms;
- The DILIGENT Information Service uses the gLite R-GMA in order to gather information about the EGEE resources, namely CEs and SEs;
- The DILIGENT Keeper service relies on the gLite Data Management feature, namely the SEs and the LFC, in order to persist in multiple copiers the DILIGENT software packages;
- The DILIGENT Content Security and Feature Extraction services make use of the gLite job management facilities in order to use the huge computing capacity provided by the Grid to perform their computational intensive tasks. E.g. the Content Security plans to use a series of Grid jobs in order to implement a complex process involving the split of the file in a number of pieces, then in a parallel and distributed fashion apply a watermarking algorithm to each piece, and finally merge all the pieces in the watermarked version of the original file;
- The DILIGENT Content Management services will interface with the gLite services providing data management facilities in order to exploit the storage capacity of a gLite based Grid infrastructure. This will allow to maintain the DILIGENT content in an appropriate number of copies, to improve the availability of the objects and the fault tolerance;
- The DILIGENT Process Management services will wrap the gLite services providing job management facilities, namely WMS, CEs, and WNs, in order to enable DILIGENT workflows to use and combine in a seamless fashion services and jobs.
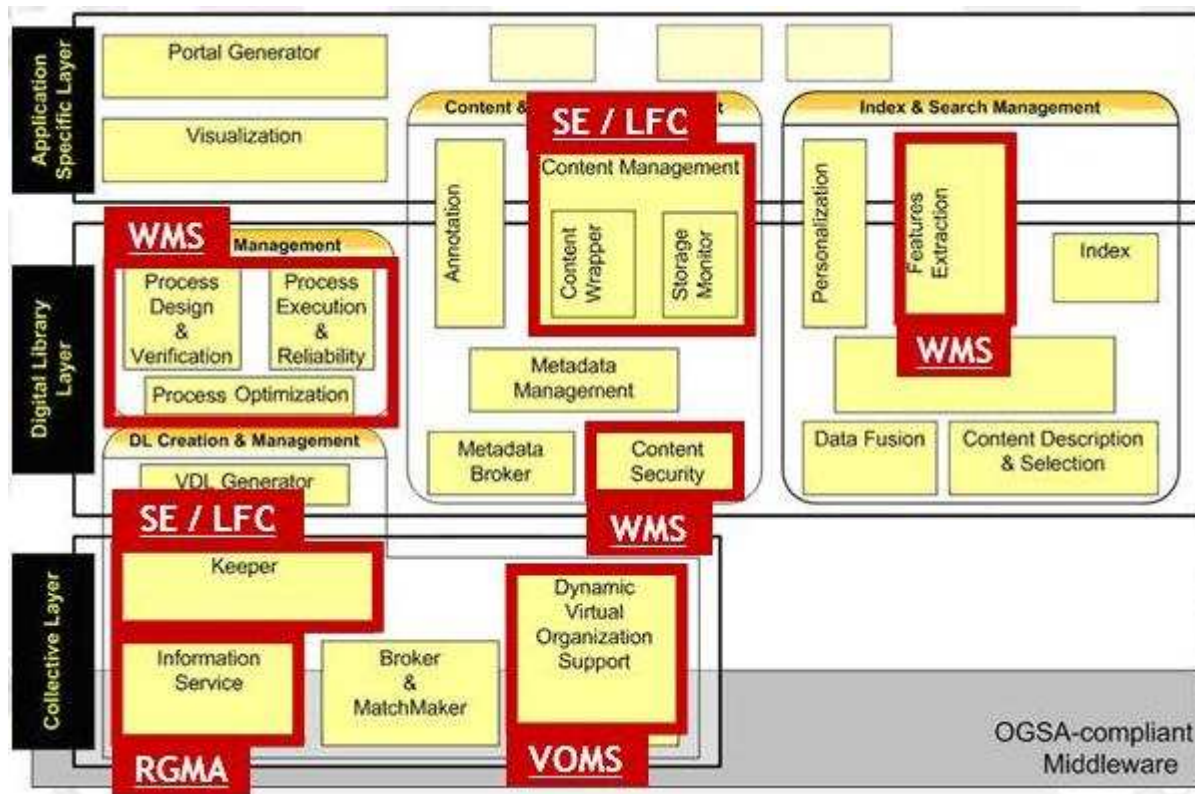
Figure 18: Mapping of DILIGENT architecture on gLite components adoption.

Infrastructure: One of the main goals of the interaction between DILIGENT and EGEE is to make available each other the resources provided. As a consequence the main dependency is related to the usage and availability of the resources constituting the various infrastructures. Each project has its plans in implementing its own infrastructure and a concrete activity is in place to integrate some pilot sites.

However the term "use" can be interpreted with different levels of engagement. The first is related to the exploitation of storage and computing capacity to maintain DILIGENT content and to process jobs respectively. The second, more demanding, is related to the capacity of the DILIGENT services to be dynamically deployed on hosting node. In this scenario the EGEE machines, if appropriately configured and equipped with the required software, can be used as nodes hosting the DILIGENT services as well.

As a consequence, the dependency of the DILIGENT infrastructure with respect to the EGEE infrastructure is related to the intrinsic characteristics of the EGEE by itself, i.e. by joining such infrastructure the DILIGENT project is enabled to exploit, under previous agreement, a huge amount of resources that otherwise it will be not able to provide. This available resources support DILIGENT in providing a level of quality of service that is difficulty to obtain by relying only on its own resources since it requires a tremendous effort in hardware resources set up and maintenance.
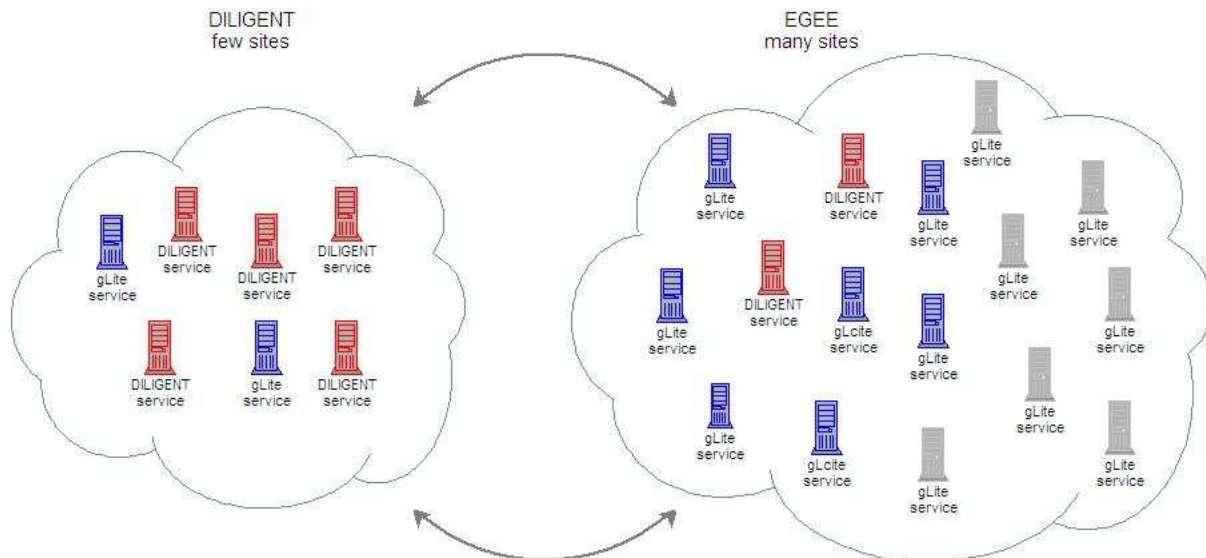
Figure 19: DILIGENT-EGEE Grid infrastructure

3.4.5   Difficulties and suggestions

The main difficulties experimented in concretely exploiting the gLite middleware are reported below.

On-going design and implementation: The adoption of a young technology, evolving over time resulted in the decision of having the design and the implementation of DILIGENT services as parallel activities, being the design an on-going process. As well, a significant amount of time had to be dedicated to experimentation, evaluation and study of the available Grid solutions. This was well known since the very beginning of the project but only in this second phase, when concrete implementation has started, it become clear that this was the best approach the project had to adopt.

Grid middleware releases adoption: EGEE has defined a precise procedure for gLite software release. Clearly some effort has to be invested in order to follow up with middleware versions. In addition, gLite 3.0 functionalities had to be studied and adopted where new components were suggested as replacement of previous versions of software components. Not differently as expected, a strategy for deployment of Grid middleware in the DILIGENT infrastructures had to be agreed and applied. EGEE representatives showed to be open and collaborative in supporting DILIGENT in this process.

Middleware requirements and bugs fixing: To have a technical requirement implemented in a large collaboration like EGEE requires good communication between the two parties (the users and the middleware team), good organisation inside the collaboration to get the requirements discussed and, when appropriate, further developments approved and implemented. All this involve necessarily a certain time for the updates to be finally included in the new software release.

Transition to EGEE II: In April 2006 EGEE entered a new phase of the project, called EGEE II. This transition didn't involved many changes for the DILIGENT project and for the already established relations with EGEE. Only few EGEE activities were significantly reorganised and the communication between the two projects (thought the TCG, NA4 and NA1 representatives) remained unchanged and effective. Nevertheless, it would have been

useful if projects like DILIGENT, already collaborating with EGEE, would have been addressed an official communication focused on the novelties that EGEE II have brought and could be of interest for their activity.

Technical communication: During the first phase of DILIGENT there was not official technical communication channel that could allow the project members to ask questions related to the adoption of gLite middleware. This was in fact implemented by setting up a number of technical joint meetings. In this second phase the TCG workgroups and the security workgroups created important opportunities for exchange, as well as the Related Projects track at the EGEE conference. DILIGENT encourages the TCG to take into account the voice of the external projects and keep them informed about the TCG resolutions and the timelines for implementation. It would also be fruitful for other EGEE collaborations if the projects were constantly informed about EGEE middleware development and support plans, as well as EGEE was informed about the status of their Grid related discussions and advancements. The communication established through the EGEE workgroups, the TCG and the EGEE technical management demonstrated to be of high value for keeping DILIGENT plans aligned with EGEE.

Documentation: Partners found documentation regarding VOMS and other gLite services sometimes unclear or not specified at the level of accuracy they would have liked. For instance, difficulties have been encountered in using gLite-voms-util API due to lack of documentation for AttributeCertificate retrieval. Only physical meetings with developers gave the opportunity to clarify open issues. Also the gLite-discuss mailing list turned out to be an effective communication channel in order to clarify on very specific issues and get specialized support on gLite installation and usage.

3.4.6   Plans for the near future

In the near future the progresses for both the development of the DILIGENT technology and the experimentation with the two user scenarios will provide a more concrete understanding of the problems related to the integration of Grid and DL technologies. This better understanding will provide important additional feedback to the Grid/EGEE and DL research communities.

At present the DILIGENT project is in a phase in which it can start to present, demonstrate and publish the first results. For example, a DILIGENT presentation was already accepted to the Global Grid Forum, while a number of papers have been submitted to the forthcoming European Conference on Digital Library (ECDL 2006) and related workshops. The dissemination of the project results at such events organised by the two communities will certainly stimulate a wider discussion which will involve others researchers and experimenters. This enlarged discussion will contribute to the realization of one of the main project objectives: "to promote cross-fertilization between the DL and Grid technology domains" in order to "foster synergies and advances in both the areas".

3.5   Final Considerations

Looking at the interactions established during this second period of reporting, a comparison with the first reporting period can be done.

The first reporting period (Period 1) refers to the first 8 months of the project lifetime and collects the interactions which took place between September 2004 and beginning of May 2005. It covers a period of 8 months.

The second report collects the interactions which took place between May 2005 and the end of May 2006 (Period 2). Therefore it covers a period of 12 months.

Period 1 reports about 26 interactions. 3.25 interactions per month in average.

Period 2 reports about 32 interactions. 2.67 interactions per month in average.

Considering an average of 18.5 working days per month this means:

- during Period 1 one interaction every 5.69 working days
- during Period 2 one interaction every 6.9 working days

Looking at the areas and the motivations for interactions reported in Table 2 of this document, the following comparisons apply.

Areas of interaction

A weighted value has been assigned to each occurrence in the table: 3 for the first, 2 for the second and 1 to the last one. The following table summarizes the results.

	Period 1			Period 2		
	#	weight	%	#	weight	%
Middleware	19	53	57	22	62	56.5
Operations	3	8	8.6	7	21	19
Networking	11	32	34.4	10	27	24.5
total	33	93	100	39	110	100

Table 12: Areas of interaction with EGEE

The following graphs are used to visually compare the collected statistics.
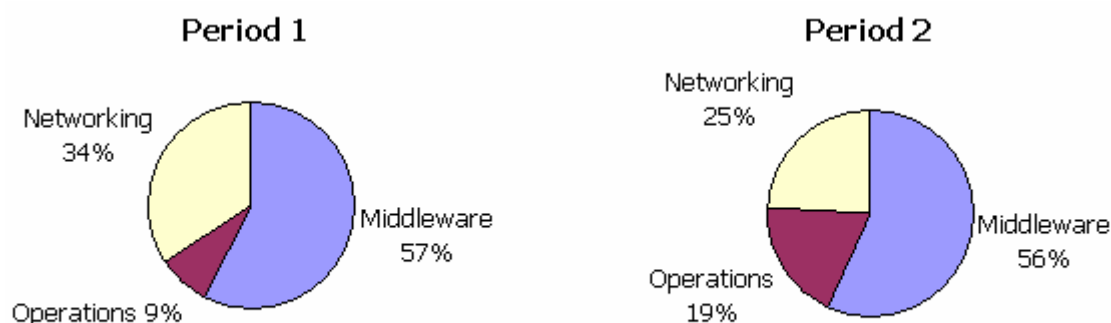

Figure 20: Areas of interaction with EGEE

Motivations for interaction

A weighted value has been assigned to each occurrence in the table: 7 for the first occurrence, 6 for the second, decreasing to 1 for the last one. The following table summarizes the results.

	Period 1			Period 2		
	#	weight	%	#	weight	%
Learn	18	124	49.8	14	80	18.3
Experiment	1	7	2.8	4	23	5.2
Provide Feedback	3	21	8.4	19	123	28.1
Deploy	3	18	7.2	5	32	7.3
Implement	3	19	7.7	12	70	16
Disseminate/Demonstrate	6	40	16.1	4	28	6.4
Align	3	20	8	13	82	18.7
total	37	249	100	71	438	100

Table 13: Motivations for interaction with EGEE

Here below, the following two graphs visually compare the collected statistics.
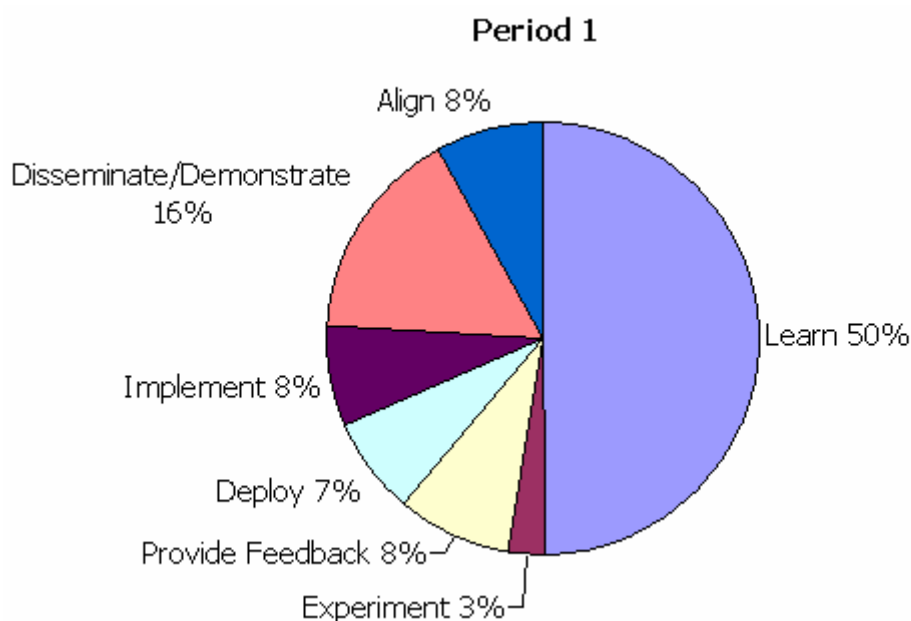

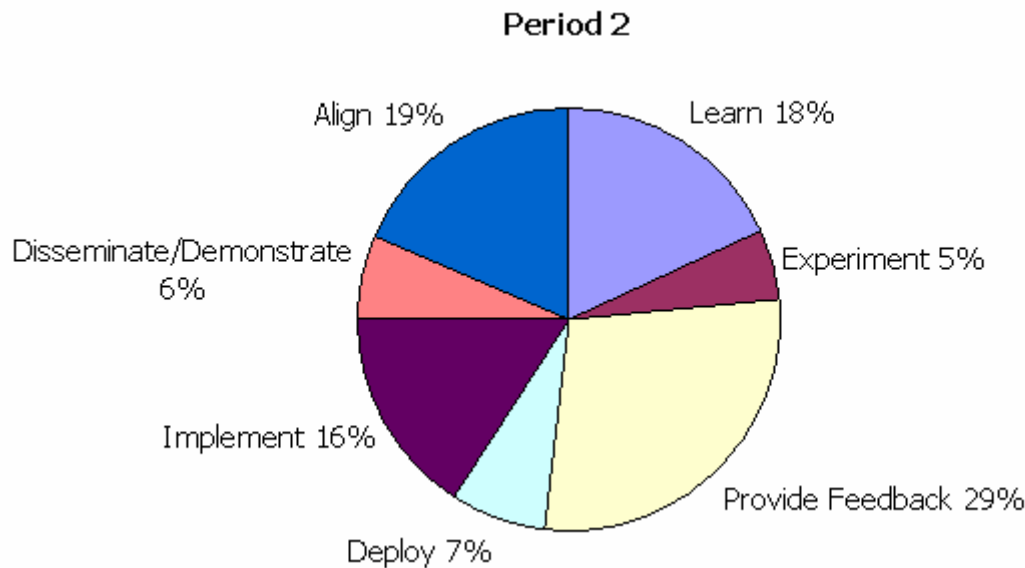*Figure 21: Motivations for interaction with EGEE related to Period 1*

Figure 22: Motivations for interaction with EGEE related to Period 2

To conclude, the graphs highlight that in both periods still a large number of interactions were related to middleware issues. The number of interactions related to the networking areas remained the same, nevertheless the percentage of interactions related to the operations area was more significant in the second period.

If 50% of the interactions was initially motivated by the need for learning about the EGEE project and software, later this motivation differentiated into the need for providing feedback and align the DILIGENT plans to EGEE plans. Still, effort was devoted to interactions related to learning and implementation activity. Here below is a classification of the motivations for interactions for the two reference periods:

Period 1		Period 2	
Learn	49.8%	Provide Feedback	28.1%
Disseminate/Demonstrate	16.1%	Align	18.7%
Provide Feedback	8.4%	Learn	18.3%
Align	8%	Implement	16%
Implement	7.7%	Deploy	7.3%
Deploy	7.2%	Disseminate/Demonstrate	6.4%
Experiment	2.8%	Experiment	5.2%

Table 14: Classification of motivations for interaction with EGEE

Finally, it has to be emphasized here that the close collaboration established during the lifetime of the two projects has extremely facilitated the adoption of the EGEE grid middleware in DILIGENT and allowed the project to keep its plans aligned with the evolving lower level young technology. DILIGENT technical partners gained a good understanding of the gLite components and provided useful feedback to EGEE. EGEE technical coordination has been opened to collaborate with DILIGENT and support it in the exploitation of gLite. The two projects have therefore taken mutual advantage from a deeper substantial technical discussion improved over time.

Appendix A.   gLite Architecture and Components

gLite is the name of the EGEE Grid Middleware. It can be seen as a collection of Grid services. gLite follows a *Service Oriented Architecture* which will facilitate interoperability among Grid services and allow easier compliance with upcoming standards, such as OGSA, that are also based on these principles.

The architecture constituted by this set of services is not bound to specific implementations of the services and although the services are expected to work together in a concerted way in order to achieve the goals of the end-user they can be deployed and used independently, allowing their exploitation in different contexts.
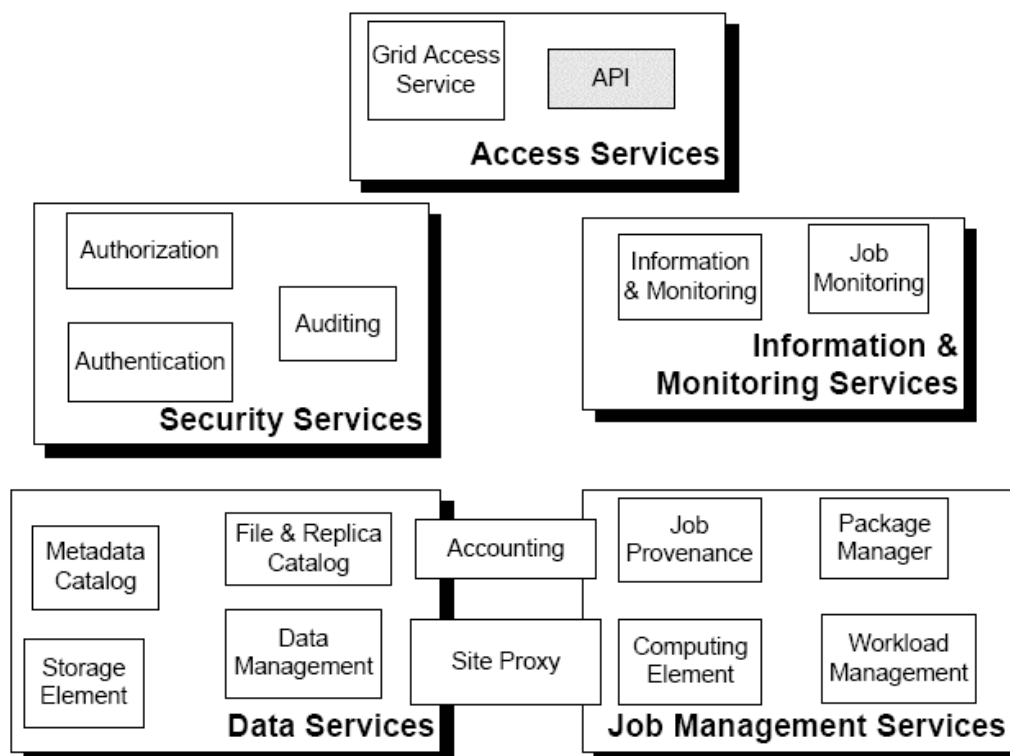

Figure 23: gLite architecture

Security services encompass the Authentication, Authorization, and Auditing services which enable the identification of entities (users, systems, and services), allow or deny access to services and resources, and provide information for post-mortem analysis of security related events.

API and Grid Access Service provides a common framework by which the user may gain access to the Grid services. The access service will manage the life-cycle of the Grid services available to a user, according to his/her privileges.

Information and Monitoring Services provide a mechanism to publish and consume information and to use it for monitoring purposes. The information and monitoring system can be used directly to publish, for example, information concerning the resources on the Grid.

Job Management Services The main services related to job management/execution are the computing element, the workload management, accounting, job provenance, and package manager services.

Data Services The three main service groups that relate to data and file access are the Storage Element, Catalog Services and Data Management.

Acronyms

ACL	Access Control Lists
API	Application Programming Interface
AuthN	Authentication
AuthZ	Authorization
CA	Certification Authority
CAS	Community Authorization Service
CE	Computing Element
CS	Compound Service
CTL	Center for the Data Processing of Texts and Images in the Literary Tradition
DL	Digital Library
DLMS	Digital Library Management System
EDG	European DataGrid Project
EGEE	Enabling Grids for E-sciencE
ESA	European Space Agency
GAS	Grid Access Service
GGF	Global Grid Forum
GSI	Grid Security Infrastructure
JDL	Job Description Language
JP	Job Provenance
JSR168	Java Specification Request: Portlet Specifications
LB	Logging and Bookeeping
LCAS	Local Centre Authorization Service
MC	Metadata Catalogue
OGSA	Open Grid Services Architecture
OS	Operating System
OWL	Web Ontology Language
PM	Package Manager
R-GMA	Relational Grid Monitoring Architecture
SAML	Security Assertion Markup Language
SE	Storage Element
SOA	Service Oriented Architecture
SOAP	Simple Object Access Protocol
UC	Use Case
VDL	Virtual Digital Library
VO	Virtual Organisation
VOMS	Virtual Organisation Membership Service
WM	Workload Manager
WMS	Workload Management System
WSDL	Web Services Description Language
WSRF,WS-RF	Web Service Resource Framework [6]
WSRP	Web Service for Remote Portlets [23]
XACML	eXtensible Access Control Markup Language