



A **D**igital **L**ibrary
Infrastructure on **G**rid
ENabled **T**echnology

Deliverable No D4.3.3a:

"Market Watch"

July 2006

Document Information

Project	
Project Title:	DILIGENT, A DI gital L ibrary I nfrastructure on Grid EN abled T echnology
Project Start:	1 Sep 2004
Call/Instrument:	FP6-2003-IST-2/IP
Contract Number:	004260
Document	
Deliverable number:	4.3.3a
Deliverable title:	Market Watch
Contractual Date of Delivery:	M22
Actual Date of Delivery:	10 October 2006
Editor(s):	Stefan Debald, Andrea Manieri
Author(s):	Pedro Andrade, Donatella Castelli, Stefan Debald, Peter Fankhauser, Veronica Guidetti, Andrea Manieri, Roberta Faggian Marque, Claudia Niederee, Markus Nikel, Silvia Volterrani
Participant(s):	CERN, CNR-ISTI, ENG, ESA, FAST, FhG/IPSI, UNIBAS, SNS, RAI
Workpackage:	WP4.3
Workpackage title:	Exploitation
Workpackage leader:	ENG
Workpackage participants:	CERN, CNR-ISTI, ENG, ESA, FAST, FhG/IPSI, UNIBAS
Est. Person-months:	
Distribution:	Internal
Nature:	Report
Version/Revision:	1.4
Draft/Final	Final
Total number of pages: (including cover)	74
File name:	D4.3.3a Market Watch_V1.4.doc/.pdf
Key words:	<i>Digital libraries, Grid, utility computing, e-Science, eu-ist</i>

Document log			
Version	Date	Comment	Author
0.1	2006-02-02	1 st TOC / Chapter Structure	Stefan Debald
0.2	2006-06-05	Contributions from CERN/ESA included	Roberta Faggian Marque, Pedro Manuel Rodrigues De Sousa Andrade, Veronica Guidetti, Stefan Debald (ed.)
0.3	2006-06-07	Initial contributions from FhG included	Peter Fankhauser, Stefan Debald (ed.)
0.4	2006-06-09	Editorial revision of first contributions	Stefan Debald (ed.)
0.5	2006-06-13	Initial contributions from ENG included	Andrea Mariani, Stefan Debald (ed.)
0.6	2006-06-14	Initial contributions from CNR included	Leonardo Candela, Donatella Castelli, Stefan Debald (ed.)
0.7	2006-06-19	Executive summary included, update report compliant with deliverable template	Stefan Debald
0.8	2006-06-22	Comments and updates from CERN/ESA, ENG, and CNR included. Template clean-up.	Roberta Faggian Marque, Pedro Manuel Rodrigues De Sousa Andrade, Andrea Manieri, Veronica Guidetti, Donatella Castelli, Stefan Debald (ed.)
0.9	2006-06-23	Feedback aggregation, harmonisation of formatting.	Stefan Debald
1.0	2006-07-12	Contribution integration, harmonisation of text and re-arrangement	Roberta Faggian Marque, Pedro Manuel Rodrigues De Sousa Andrade, Veronica Guidetti, Donatella Castelli, Stefan Debald, Andrea Manieri (ed.)
1.1	2006-07-19	Inclusion of user community analysis	Markus Nickel, Silvia Volterrani, Stefan Debald (ed.)
1.2	2006-07-23	Update of conclusion and finalisation of deliverable	Stefan Debald
1.3	2006-08-01	Consideration of reviewer comments	Roberta Faggian Marque, Markus Nickel, Veronica Guidetti,

			Stefan Debald (ed.)
--	--	--	---------------------

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 contents of this publication are 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

1 EXECUTIVE SUMMARY

The objective of this Market Watch report is to communicate findings of the monitoring regarding the evolution of the Digital Library (DL) and Grid markets, to update the project on current trends (technological as well as marketwise), related products, and emerging competitors. In addition, this report will support the definition and refinement of the business model for DILIGENT which will be carried out in D4.3.4a "Preliminary Business Plan".

The DILIGENT project targets the development of a DL infrastructure testbed which facilitates knowledge sharing and remote co-operation in e-Science. The DILIGENT infrastructure will support the on-demand creation of a new generation of DL which will be built by exploiting a given set of shared resources maintained by the infrastructure itself. These shared resources, which may be processing power, storage, instruments, data or applications, will be managed dynamically and their allocation will be a transparent process.

The DILIGENT solution **is a framework for the dynamic on-demand creation of DL for e-Science communities and organizations based on a Grid infrastructure.** In this vision DILIGENT **is creating a new value chain in the DL business models.** The framework allows Virtual Research Organisations (VRO) to integrate archives and other services required to meet VRO needs. The idea to bring together Grid and DL technologies will enhance not only these technologies per se, but also define systematic ways of how to develop next generation information networks on the basis of Grid technologies, which can be applied and adapted to other domains.

This Market Watch report is designed to be an on-going deliverables that will be improved in two distinct iterations during the project lifetime. This first edition of the report focuses on a revision of the market analysis carried out in D4.3.1 "Market and Technology Trends Analysis". Due to the updating character of this report it will structurally be close to D4.3.1. In addition to an update of the DL domain as described previously, a direct functional comparison of existing DL and repository solutions is carried out. Furthermore, for the first time, this report includes the investigation of existing best practices in the user communities of DILIGENT and the disruptive potential of expected DILIGENT results through analysis of the possible impact on the user communities.

In addition to this executive summary (**Chapter 1**), this document contains eight chapters that carry out the analysis as described above and will give input to the definition of the DILIGENT business plan. An initial chapter (**Chapter 2**) gives a short description of the DILIGENT project and introduces our analysis methodology. **Chapter 3** updates the analysis of the functional perspective focussing on Digital Library and Scientific Repositories domains. A case-based direct functional comparison of well established applications from the different domains is carried out. In addition, in the context of scientific repositories, recent studies which try to find a new model for scientific publication are discussed. **Chapter 4** introduces the user communities' point of view in the analysis of the expected impact of DILIGENT. As an initial step of this investigation the representatives of the user communities which are also partners in DILIGENT have been interviewed on their present best practices and anticipated impact of the DILIGENT results. **Chapter 5** gives an update of the grid as the primary enabling technology for DILIGENT, including an overview of the grid market and current standardisation trends. Particular attention is given to the intensifying collaboration between DILIGENT and EGEE. Emerging scenarios in digital content management are described in **Chapter 6**. These trends are particularly important to DILIGENT as the digital content market will be the main reference market for distributing or using the DILIGENT results. **Chapter 7**

reassesses internal and external factors which are critical to the success of DILIGENT. This revision includes an update of the SWOT analysis. Finally, chapter (**Chapter 8**) summarizes the results and drawn conclusions. Appendices are contained in **Chapter 9**.

Table of Contents

Document Information	2
Disclaimer	5
1 Executive Summary	6
Table of Contents	8
2 Background	10
2.1 DILIGENT in a nutshell	10
2.2 Methodology	11
3 The functional perspectives	12
3.1 Digital Libraries evolution	12
3.1.1 Repository Systems	13
3.1.2 Digital Library Systems	17
3.2 Scientific Repositories perspectives	22
3.3 Outcomes	24
4 Impact on Communities	25
4.1 Earth Observation Science: the ESA case	25
4.1.1 Community characterization	25
4.1.2 Best practices	26
4.1.3 The expected impact of DILIGENT	26
4.2 Humanities: the SNS experience	27
4.2.1 Community characterization	28
4.2.2 Best practices	28
4.2.3 The expected impact of DILIGENT	29
4.3 Content Provider: the RAI point of view	29
4.3.1 Community characterization	30
4.3.2 Best practices	30
4.3.3 The expected impact of DILIGENT	30
4.4 Outcomes	30
5 Grid as enabling technology	32
5.1 Overview of the market and trends	32
5.2 gLite vs. other solutions	33
5.2.1 Other Grid Middleware Initiatives	34
5.2.2 Grid Standards and their evolution	36
5.3 Outcomes	37
6 Emerging Scenarios in Digital Content Management	38
6.1 Pervasive Media: the Content market in 2010	38
6.2 What content for the Digital Content Market	39
6.3 Openness vs. Rights protection	41
6.4 Outcomes	43
7 Reassessment of internal and external factors	44
7.1 Critical success factors	44
7.2 SWOT Analysis	44
8 Conclusions	46

9	Appendices	49
9.1	Focus on EGEE	49
9.2	Other Grid middleware initiatives	57
9.3	Grid standardisation bodies	65
9.3.1	Distributed Management Task Force (DMTF)	65
9.3.2	Internet Engineering Task Force (IETF)	65
9.3.3	Global Grid Forum (GGF)	66
9.3.4	Organization for the Advancement of Structured Information Standards (OASIS)	67
9.3.5	Web Services Interoperability (WS-I)	67
9.3.6	World Wide Web Consortium (W3C)	67
9.3.7	Others	67
9.4	Existing Standards in grid computing	68
9.4.1	Extensible Markup Language (XML)	68
9.4.2	Grid Security Infrastructure (GSI)	68
9.4.3	Lightweight Directory Access Protocol (LDAP)	68
9.4.4	Open Grid Services Architecture (OGSA)	69
9.4.5	Secure Socket Layer (SSL)	69
9.4.6	Simple Object Access Protocol (SOAP)	69
9.4.7	Universal Description, Discovery and Integration (UDDI)	69
9.4.8	Web Services Description Language (WSDL)	69
9.4.9	Web Services Interoperability Basic Profile (WSI BP)	70
9.4.10	Web Services Addressing (WS-Addressing)	70
9.4.11	Web Services Management (WS-Management)	70
9.4.12	Web Services Notification (WS-N)	71
9.4.13	Web Services Reliability (WS-Reliability)	71
9.4.14	Web Services for Remote Portlets (WSRP)	71
9.4.15	Web Services Resource Framework (WSRF)	71
9.4.16	Web Services Security (WS-Security)	72
10	Bibliography and References	73

2 BACKGROUND

This is an introductory chapter whose aim is to provide basic knowledge to the reader regarding DILIGENT project, and the methodology adopted in performing this study.

This chapter is similar to the analogous of D4.3.1 Technology and Market trends analysis with some updates due to the past experience. In particular the methodology is adapted to focus more on the market and business aspects as well as on the impact to end-user communities.

2.1 DILIGENT in a nutshell

Digital Libraries have been mainly intended as the digital counterpart to physical libraries. This understanding has been maintained for many years, even when Digital Libraries have been proved to be applicable not only to the library domain but to the entire Cultural field. Recently Digital Libraries have moved beyond the traditional connotation of the term "library", and are rapidly shifting towards more general systems, now termed "dynamic universal knowledge environments" [17].

These new Digital Libraries will be able to support the work of its users by providing functionalities that may range from general utilities, like annotation, summarization or co-operative work support, to user specific functions, like processing of maps, semantic analysis of images, simulation, etc. Through these new Digital Libraries, groups of individuals, which collaborate together to achieve a common goal, will be allowed to access, discuss and enhance the on-line shared information produced by them and by other groups or individuals.

The objective of the DILIGENT project is to develop a digital library infrastructure that facilitates knowledge sharing and remote co-operation in e-Science. The distributed and dynamic nature of scientific collaborations requires infrastructures and applications to be able to adapt to the needs of various user communities. Thus, this infrastructure will support the on-demand creation of new generation Digital Libraries that will automatically be built by exploiting the set of shared resources maintained by the infrastructure itself. These resources are **content sources** (i.e., repositories of information searchable and accessible through a single "entrance"), **services** (i.e., software tools, that implement a specific functionality and whose descriptions, interfaces and bindings are defined and publicly available) and **hosting nodes** (i.e., networked entities that offer computing and storage capabilities and supply an environment for hosting content sources and services). By exploiting the functionality offered by this infrastructure multiple Virtual Research Organisations will be allowed to share their resources according to established policies and to build Digital Libraries that satisfy their specific co-operation needs.

Further, the increasing demand for intensive computation and processing of very large amounts of information (data, multimedia, documents, etc.) highlights the need to perform experiments at lower costs. In this context Grid computing has gained a lot of attention within the academic and scientific community and the IT industry.

The DILIGENT infrastructure is being built by integrating Digital Libraries (DLs) and Grids technologies. The idea to intertwine Grid and DLs not only enhances these technologies per se, but also defines systematic ways on how to develop next generation information networks on the basis of Grid technologies, which can be applied and adapted to other domains. The underlying idea is to promote the future generation of technologies in which computers and networks will be integrated into the everyday environment, rendering accessible a multitude of services and applications through easy-to-use human interface. In the Grid vision, once a proper kind of infrastructure is in place, a user will have access to a "virtual computer" that is reliable and adaptable to the user's needs. The resources –

which may be processing power, storage, instruments, data or applications – will be managed dynamically and their allocation will be a transparent process.

The DILIGENT solution is thus a framework for dynamic creation and maintenance of Digital Libraries for e-Science communities and organizations. Using this framework any Virtual Research Organisation will be able to integrate any kind of archive and any third party services that could complement basic functions to accomplish with VRO needs.

DILIGENT organization will supply a service for the creation of on-demand Dynamic Digital Libraries in a Pay-per-use fashion, putting together resources eventually supplied by third parties. The DILIGENT service is related either with the creation of the Digital Libraries based on VRO needs and with the maintenance (and potentially the evolution) during the Digital Libraries lifetime.

In this vision DILIGENT creates a new value chain in the Digital Libraries business model, bringing together Content Providers, Service Providers and Storage and/or Computation Providers, to supply a new service to Institutions and Organisations (mainly virtual) that need to create Digital Libraries for their end-users. Examples of DILIGENT customers are, hence, public and private libraries, pools of federated libraries (maybe local libraries), temporary aggregation of institutions (e.g., crisis management teams, events organisation teams, etc), conference organisers, etc.

2.2 Methodology

This Market Watch analysis has been made during the second year of the project and after an already performed "Technology and Market trends Analysis" D4.3.1.

While the latter had the aim to conduct a multi-level analysis on both enabling technologies and market trends, this one will concentrate on the Market perspective and user impact.

The reason is that during the first year the project had to be assessed both in term of technologies and market expectations. Currently the technology chosen (grid technology) is well established and with promising stable future, while the Market is rapidly changing.

In particular, the analysis aims to understand the international market in which the project will compete, focusing on the following aspects:

1. the **first aspect refers to the organisational perspectives and sustainable models** identifying analogous experience in Digital Libraries and initiatives supporting Virtual Communities;
2. the **second issue refers to grid technologies evolution and standardization perspectives**, since it is the main technology on which DILIGENT is leveraging and that may considerably affect future technical choices;
3. **the third issue is related to the trends in the Digital Convergence and the appearance of new value chains**
4. Finally, the **market composition** applies all the considerations, technology and market insights, information and knowledge extracted, to identify the possible market addressable by DILIGENT, both in terms of services to offer and users or customers needs.

3 THE FUNCTIONAL PERSPECTIVES

Digital Libraries constitute a relatively young scientific field, whose life spans roughly the last fifteen years. Instrumental in the birth and growth of the field have been the funding opportunities generated by the 'Technology Enhanced Learning; Cultural Heritage' (formerly 'Cultural Heritage Applications') Unit of the Information Society Directorate-General of the European Commission and the 'Digital Library Initiatives' in the United States sponsored by the National Science Foundation and other agencies.

In Europe in the last years began critical the adoption of some strategies to manage the huge amount of data (mainly of scientific nature) developed within the IST and other EC-funded programs. Recently a workshop has been set on the theme of Scientific Repositories to understand the state-of-the-art and trace the further evolution of this field.

In this chapter we briefly outline the current trends on these two areas (Digital Libraries and Scientific Repositories) as benchmarking of DILIGENT capabilities to supply all the communities from both sides.

3.1 Digital Libraries evolution

Digital Libraries represent the meeting point of a large number of disciplines and fields, i.e., data management, information retrieval, library sciences, document management, information systems, the web, image processing, artificial intelligence, human-computer interaction, and others. It was only natural that these first fifteen years were mostly spent on bridging some of the gaps between the disciplines (and the scientists serving each one), improvising on what "Digital-Library functionality" is supposed to be, and integrating solutions from each separate field into systems to support such functionality, sometimes the solutions being induced by novel requirements of Digital Libraries. These have been achieved through much exploratory work, primarily in the context of focused efforts devising specialized approaches to address particular aspects of Digital-Library functionality. For the most part, every effort so far has been distinct and, in some sense, isolated from the rest. Every project has started from scratch to build a system supporting the particular needs specified in the project's description. Nevertheless, looking back at the individual achievements of all the projects, one may see clearly that there is substantial commonality among many of them; the bottom-up development of the field so far has provided enough 'data points' for patterns to emerge that can encapsulate all efforts.

Despite the young age of the field of Digital Libraries, it has made a long journey from its initial conception to the present state of the art and has reached a level of maturity that did not exist fifteen years ago. There is substantial knowledge and experience that have been accumulated.

The concept of Digital Library itself has evolved quite substantially since the early idea of a system providing access to digitized books and other text documents. The DELOS Network of Excellence on Digital Libraries **Error! Reference source not found.** now envisions a Digital Library as a tool at the centre of intellectual activity having no logical conceptual, physical, temporal, or personal borders or barriers on information. It has moved from a content-centric system that simply organizes and provides access to particular collections of data and information, to a *person-centric system* that aims to provide interesting, novel, personalized experiences to users. Its main role has moved from static storage and retrieval of information to facilitation of communication, collaboration, and other interaction among scientists, researchers, or the general public on themes that are pertinent to the information stored in the Digital Library. Finally, it has

moved from handling mostly centrally-located text to synthesizing distributed multimedia document collections, sensor data, mobile information, pervasive computing services, etc.

This vision of Digital Libraries seems to resonate well with the concept of "Information Space" that has arisen from the field of Computer Supported Cooperative Work (CSCW). Snowdon, Churchill, and Frecon [3] have developed future visions about "Connected Communities" and "Inhabited Information Spaces", with the latter being closely related with the vision of Digital Libraries, in that ubiquitous information is a prerequisite for CSCW. In more detail, Inhabited Information Spaces are "spaces and places where people and digital data can meet in fruitful exchange, i.e., they are effective social workspaces where digital information can be created, explored, manipulated and exchanged". Thus, "in Inhabited Information Spaces, both information and people who are using that information (viewing it, manipulating it) are represented. This supports collaborative action on objects, provides awareness of others' ongoing activities, and offers a view of information in the context of its use". Based on the above and according to the vision of a Digital Library, the latter provides an Information Space that is populated by a user community and becomes an Inhabited Information Space through CSCW technology. The two fields complement each other nicely, in that one focuses on access and provision of relevant information while the other focuses on visualisation and sharing of information.

3.1.1 Repository Systems

An increasing range of activity areas within the DL environment refers to their deposited content collections as "repositories". Anderson and Heery's examination of repository models provides an excellent overview of the characteristics of repositories, although it does not delve into any repositories very deeply [4]. In the following we discussed very briefly three well known repository applications, DSpace, Fedora, and aDORe.

DSpace [5] is an open source digital repository software system for research institutions. It has been developed jointly by the MIT Libraries and Hewlett-Packard Labs, and it is available under the BSD open source license for research institutions to run as-is, or to modify and extend as needed. It enables organizations to:

- Capture and describe digital material using a submission workflow module, or a variety of programmatic ingest options
- Distribute an organization's digital assets over the web through a search and retrieval system
- Preserve digital assets over the long term.

DSpace reports over 80 installations reported mainly in publicly funded institutions, which has not increased much since last year. It is coordinated by the DSpace Federation, an informal collection of active DSpace users, originally funded (2003-2004) by the Andrew W. Mellon Foundation, which also holds user conferences.

Fedora [6] (Flexible Extensible Digital Object and Repository Architecture) supports interoperability and extensibility of digital library systems and institutional repositories. The Digital Library Research Group at Cornell University originally developed the Flexible Extensible Digital Object Repository Architecture (Fedora) under a National Science Foundation Grant. The transition of Fedora from a research prototype to production repository software began when the University of Virginia Library, seeking a solution for managing increasingly complex digital content, experimented with the Fedora architecture. The experimentation proved successful, providing the basis for subsequent funding from the Andrew W. Mellon Foundation to Cornell and Virginia to jointly develop Fedora and make it available as open source software to libraries, museums, archives, and content managers, facing increasing variety and complexity in the digital content that they manage. Mellon-funded development continues through 2007.

The software is available under the terms of the Educational Community License 1.0 (ECL) Fedora reports about 35 installations mainly in publicly funded institutions, which has also not increased much since the last year. Industrial Vendor Support is carried out by VTLS (Visionary Technology in Library Solutions), however their website (www.vtls.com) does not contain any indication of actual industrial fedora installations. For sustaining Fedora beyond its funding, an advisory board has been created.

aDORe [7] is a repository system designed and implemented at Los Alamos National Laboratory for enabling self-supporting access to digital information. It hosts the vast collection of digital scholarly assets that the LANL Research Library acquires or licenses (approximately 80,000,000 on June 2005) and makes them accessible through locally developed user services. In addition, it has been publicly released and is distributed as open source under LPGL (GNU Lesser General Public License). As it has been only recently released (May 17, 2006) no (external) usages are reported.

The table below, extracted from [8], shows the functionalities of the three repository systems when analyzed with reference to the DELOS Reference Model for Digital Libraries [9]. Those aspects for which we have found no description in the literature have been left unspecified.

	Fedora	DSpace	aDORe
User			
User Identifier	Yes	Yes	No
User Profile	No	Yes	No
Role	No	Yes	No
Policy	No		No
Group	No	Communities	No
Information Space			
Information Object	FOXML (supports METS, too)	Relational Database	MPEG-21 DID
o Information Object Identifier	FEDORA Persistent Identifier (PID)	CNR Handle System	Identifier Locator
o Content			
• Metadata	Yes	Yes	Yes
• Text	Yes	Yes	Yes
• Image	Yes	Yes	Yes
• Audio	Yes	Yes	Yes
• Video	Yes	Yes	Yes
• Composite	Yes	Yes	Yes
o Version	Versioning of individual datastreams (linear)	ABC Harmony Datamodel	Yes
o Manifestation	datastream	bitstream	datastream
o Annotation			
o Metadata			MPEG21 DIDL
• Descriptive Metadata Format	DC (searchable), any other format non-searchable	Qualified DC	

• Structural Metadata Format	FOXML	Relational Database	
• Administrative Metadata Format	FOXML	Relational Database	
• Preservation Metadata Format	FOXML, Functional preservation is implemented by the user through the	Relational Database, Bundles and Bitstreams. Functional preservation through supported file formats, bitstream preservation otherwise	
Collection	RDF based Object-to-object relationships	Native support for Collections in the Relational schema	OAI-PMH sets
Functionality			
Access			OAI-PMH
o Search	Yes	Yes	No
• Full Text	No	No	No
• Metadata	Yes	Yes	No
• Image	No	No	No
• Audio	No	No	No
• Video	No	No	No
• Speech	No	No	No
• Single-Object, Single-Feature	No	No	No
• Multi-Object, Multi-Feature	No	No	No
• Compound Document Match	No	No	No
• Predicates	No	No	No
• Query Expansion	No	No	No
o Cross-language	UTF-8	UTF-8	No
o Relevance Feedback	No	No	No
o Browse	Very Basic	Yes, built-in support in the Web UI	No
o Visualize	Very Basic	Yes, built-in support in the Web UI	No
o Translate	No	No	No
Content Management			
o Submit	Administrator only, perhaps in FEDORA 2.1 updated Authorization model other roles may exist too	Yes	
o Update	Administrator only	Yes	
o Annotate			
o Review	Administrator only	Yes	

DL Management			
o Annotate			
o Update	Yes	Yes	
o Withdraw	Yes	Yes	
o Describe	Yes	Yes	
o Disseminate	Yes	Yes	
o Preserve	Yes	Yes	
o User Management			
• Registration	No	Yes	
• Management Role	No	Yes	
o Policy Management		Yes	
Personalize			
o Collection Management	No	Yes but basic	No
o Personalised access	No	Yes	No
o Notification	No	Yes	No
o Others		Online registration	No
Enabling			
o Authentication	Yes	Yes	
o Authorization	Yes (XACML-based Policy Enforcement)	Yes (Relational Database)	
o Encryption	No	No	
o Subscription	No	Yes	
o Notification	No	Yes	
o Process composition	No, only dissemination of behaviors on datastreams	Yes (simple 3-step workflow)	
Others	<ul style="list-style-type: none"> OAI-PMH harvesting Most functions are exposed as SOAP-based Web Services Support for LDAP, IP-based authentication, HTTP basic authentication and SSL-based authentication 	<ul style="list-style-type: none"> OAI-PMH harvesting Support for LDAP 	<ul style="list-style-type: none"> OAI-PMH Federator Open URL Resolver
Quality of Service			
Security			
Economics			
Availability			
Reliability			
Performance			
o Response time			
Security			
o Authentication			

o Integrity			
o Data Protection			
o Message Protection			
Robustness			
Capacity			
Load balancing			
Recoverability			
Messaging			
Consistency			
Scalability			
Architecture			
Characteristics	Loosely coupled services acting on top of the repository service.	Layered architecture of components interacting through API.	Component-based and standard-based (XML, MPEG21 DID and DII, OpenURL, OAI-PMH) architecture where interaction is protocol-based.

Table 1: Functional comparison of repository systems

3.1.2 Digital Library Systems

Digital Library Systems are different from repository systems because their goal is to provide a broader range of functionality than is provided by repository systems. They are of a general purpose nature, have been implemented to fulfill the requirements of particular types of DL building user communities. In this section we examine OpenDLib, OSIRIS/ISIS, and Daffodil, three digital library systems developed by DELOS partner institutions and discuss their capabilities and facilities for building digital libraries.

OpenDLib [10] is a software toolkit developed at ISTI-CNR that can be used to easily create a digital library, according to the requirements of a given user community. This can be done by first instantiating the software appropriately and then either loading or harvesting the content to be managed. The toolkit consists of a federation of services that implement the digital library functionality making few assumptions on the nature of the information objects to be stored and disseminated. Using the toolkit it is possible to handle a wide variety of information object types with different formats, media and structures. In particular, the toolkit can manage new types of information objects that have no physical counterpart, such as composite information objects consisting of slides, video and audio recordings of lectures, seminars or courses. OpenDLib can also maintain multiple editions, versions, and manifestations of the same information object, each described by one or more metadata records in different formats. The information objects can then be organized in a set of virtual collections, each characterized by its own access policies. Authorized people can dynamically define new collections by specifying appropriate definition, can share private content with other selected users, and can access the digital library management functionality. The basic release of OpenDLib provides services to support the submission, description, indexing, search, browsing, retrieval, access, preservation and visualization of information objects. OpenDLib is currently not available as open source, but is used in a number of projects to create and maintain digital libraries.

OSIRIS/ISIS. OSIRIS (Open Service Infrastructure for Reliable and Integrated process Support) [11], is a platform that allows combining different distributed services into processes. The OSIRIS platform itself does not provide any application functionality but, by combining specialized application services, supports the definition and reliable execution of dedicated processes (this is also known as “programming-in-the-large”). ISIS stands for Interactive Similarity Search and is an application for information retrieval in multimedia collections built at ETH Zürich [12]. It supports content-based retrieval of images, audio and video content, and the combination of any of these media types with sophisticated text retrieval.

Daffodil [13] is a virtual DL targeted at strategic support of users during the information search process. For searching, exploring, and managing DL objects, Daffodil provides information seeking patterns that can be customized by the user for searching over a federation of heterogeneous digital libraries. Searching with Daffodil makes a broad range of information sources easily accessible and enables quick access to a rich information space.

Daffodil has been developed at the University of Dortmund, Germany and is now maintained at the University of Essen/Duisburg, Germany. Its particular focus is on supporting advanced, search strategies. Daffodil is not available as open source software, but is hosted by the Daffodil team for registered users. Since last year, the number of registered users (100) has not increased. The underlying business model appears to be mainly a playground for academic research.

The table below, again extracted from [8], summarizes the features of the DL systems described above.

	OpenDLib	OSIRIS/ISIS	DAFFODIL
User			
User Identifier	Yes	Partial (login Username / Password)	Yes (login Username/Password)
User Profile	Customisable.		Customisable.
Role	No ¹	User / Admin	User
Policy	Yes ²		
Group	Yes	Yes	Yes
Information Space			
Information Object	Compliant with DoMDL.		XML BibTex.
o Information Object Identifier	Yes	Yes	Yes
o Content			
• Metadata	Yes	Yes	Yes
• Text	Yes	Yes	Yes
• Image	Yes	Yes	No

¹ Not explicitly supported. Policies assigned per User/Group.

² Resources subject to policies are groups, collections, information objects, and services. Default actions are *create*, *edit*, *delete*, *access*, and *manage*.

• Audio	Yes	Yes	No
• Video	Yes	Yes	No
• Composite	Yes	Yes	No
Version	Internal.	Partial (No full version control, but date of last update)	
Manifestation	Multiple manifestations per object, multiple media formats.		
Annotation	No	No	Yes
Metadata			
o Descriptive Metadata Format	Yes ³		
o Structural Metadata Format	Yes ⁴	Yes	Yes, XML
o Administrative Metadata Format	Yes ⁵		
o Preservation Metadata Format	No		
Collection	Yes	ETHWorld: 625,000 images extracted from ETH websites plus corresponding textual information, ISIS: 53,837 images plus corresponding textual information, ISIS Video: 1,200 video sequences from five movies plus gathered textual meta information (cast, taglines, subtitles, keywords...), ISIS Audio: 1,185 MP3 music files plus gathered textual meta information (artist, title, album, lyrics...), ISIS Med: 50,143 medical images plus textual annotations	computer science sources: Achilles, DBLP, LeaBib, Citeseer, ACM, HCIBib, SpringerLink, CompuScience, Scirus
Functionality			
Access			
o Search	Simple and advanced. Fields set customisable.	Yes (Text + Multimedia Part)	Yes (Structured text based)
• Full Text	Yes	Yes	
• Metadata	Yes	Yes	Yes
• Image	No	Yes	No
• Audio	No	Yes	No
• Video	No	Yes	No

³ Any descriptive metadata format can be managed.

⁴ Used to represent DoMDL documents and thus compliant with a proprietary XML schema.

⁵ Proprietary format.

• Speech	No		No
• Single-Object, Single-Feature	Yes	Yes	
• Multi-Object, Multi-Feature	No	Yes	
• Compound Document Match	Yes	Yes	
• Predicates		Yes	
• Query Expansion	No		
○ Cross-language	Yes	No	UTF-8
○ Relevance Feedback	Yes	Yes	Implicit
○ Browse	Yes. Fields set customisable.	Yes	Yes
○ Visualize	Window based and tab based.	Yes (Ranked List, Fastmap)	Yes
○ Translate	No	No	No
Content Management			Done by DL sources
○ Submit	Yes	Yes	
○ Update	Yes	Yes (automatic crawling)	
○ Annotate	No		Yes
○ Review	Yes		
DL Management			Done By adding a new wrapper.
○ Annotate	No	No (but process-supported information enrichment)	
○ Update	Yes	Yes	
○ Withdraw	Yes	Yes	
○ Describe	Yes	Yes	
○ Disseminate	No		
○ Preserve	No		
○ User Management			
• Registration	Yes	Yes	
• Role Management	No	Yes	
○ Policy Management		Yes	
Personalize			
○ Collection Management		Yes	Yes
○ Personalised access		Partially (supports different namespaces; for each namespace, a new configuration can be defined and the templates for displaying the results can be exchanged)	Yes
○ Notification	No	Yes	Yes

o Others			Annotations
Enabling			
o Authentication	Login via user name and password	Yes (by username / password)	
o Authorization	Yes	Partially (by username / password; yet it is not possible to block one single service for one user or group)	
o Encryption	No		
o Subscription	No	Yes	
o Notification	No	Yes	
o Process composition	No	Yes	
Others			
Quality of Service			
Security			
Economics			
Availability	Yes	Yes (replication of services)	
Reliability	Yes	Yes (sophisticated failure handling)	
Performance	Yes	Yes	
Response time		Yes (good response time behaviour, even for highly complex similarity queries with several reference objects in collections with > 600,000 documents, response time is acceptable)	
Security			
Authentication	Yes		
Integrity		Yes	
Data Protection			
Message Protection			
Robustness	Yes	Yes	
Capacity	Yes	Yes	
Load balancing	Yes	Yes	Yes
Recoverability		Yes	
Messaging			
Consistency		Yes	
Scalability	Yes	Yes	
Architecture			
Which kind of architecture? Which are the main components and their functionality? See the reference architecture for the description and the	OpenDLib is a federation of services. These services	P2P Workflow Execution (OSIRIS)	Agent-based service oriented architecture, with backend-services and full

terminology	cooperate (i) through the OLP protocol that regulates the exchange of information about services status and (ii) with the support of the Manager service that gathers these information and provides a picture of the whole federation, checks their consistency, and controls the flow of communication.		application, java based graphical frontend.
-------------	---	--	---

Table 2: Functional comparison of digital library systems

3.2 Scientific Repositories perspectives

The European Commission has finally released its report on scientific publishing and now has firmly placed itself in the international discussion of where such publishing should go in the future. In June 2004, the European Commission began a study to examine the economic and technical evolution of scientific publishing in Europe. <http://europa.eu.int/comm/research/science-society/pdf/scientific-publication-study_en.pdf>.

The intent of this study was to find a model for scientific publication that could involve readers, authors, publishers, and funding bodies.

The study acknowledged that much of the scientific research conducted in Europe is publicly funded and hence recommended that access to such research should be guaranteed. It also contains numerous recommendations; some of them are noteworthy because they validate other proposals or offer new insights into mechanisms for guiding the future of scholarly research.

The first recommendation is: "Guarantee public access to publicly-funded research shortly after publication." Note that the following actions could be taken at the European level: "Establish a European policy mandating published articles arising from EC-funded research to be available after a given time period in open access archives." Secondly, explore with Member States and with European research and academic associations whether and how such policies and open repositories could be implemented.

Next, it is encouraged that there be a "level-playing field" so that different business models in publishing can compete fairly in the market. "It seems desirable to allow for experimentation and competition between various possible business models." The report noted that money should be allocated to libraries to subscribe to reader or library-pay journals "but also to authors to pay for publication costs in author-pay journals, and to researchers in the reader-pay model."

It is also recommended that ranking of journal quality be raised beyond "scientific quality, stricto sensu." While citation counts should remain the dominant criterion, "dimensions

related to the quality of dissemination (self-archiving authorization, publisher archiving provisions, copyright provisions, abstracting and indexing services, reference linking, etc.) could be tracked explicitly and possibly valued by research funding bodies.”

Another recommendation is to guarantee perennial access to scholarly journal digital archives by promoting the creation of “not-for-profit long-term preservation archives which balance interests among publishers, libraries, and scholars.” To this end, the report encourages an investigation into the feasibility/desirability of creating a European preservation organization that would be “JSTOR-like⁶.”

The study recommended the promotion of pro-competitive pricing strategies, noting that the limited savings that libraries obtain for cancelling subscriptions “does make it hard for newcomers to have access to library budgets.” It is suggested that simple rules could be followed. For example, the “price of electronic access should not depend on the historical number of print subscriptions” but instead should be related to “transparent indicators, like usage or the number of faculty [and] students, as is the case for JSTOR.”

In addition, the development of electronic publications should be promoted by eliminating the “unfavourable tax treatment of electronic publications” by either reducing the VAT rate or introducing a tax refund. The differences in VAT rates applied to print versus electronic journals “induced a bias in the libraries’ decisions to continue subscribing to print journals, along with the electronic version.” The authors noted that the “higher rate applied to electronic delivery of information in Europe strongly affects European research institutions, especially when compared to other countries where electronic services are exempt from tax.” Furthermore, public funding and public-private partnerships should be formed to create journal digital archives in areas when there is little commercial interest, such as in the social sciences and the humanities.

The report also strongly favoured the development of open access archives, noting that they provide “immediate, free, and maximal access to research results, whether published or not, to anyone with an Internet connection.” And, those institutional repositories contribute to “raise the profile of the institutions, making their research output visible and accessible, and provide a potential research assessment tool.” In turn, this enhanced visibility and accessibility “may lead to higher citation,” noting that recent studies show that open access increases impact. However, there are concerns about the archival quality of the open access archives. Observing that the installation costs are low, the “maintenance costs are more difficult to plan, as they will vary with the number of records, and the long term preservation purposes.”

Specific actions at the European level to improve visibility include “establish[ing] a European policy mandating articles funded from European sources to be available in open access archives, for instance by mean of author’s self-archiving.” Also, there is a need to “specify standards that will insure that the archives are [accessible], interoperable, and have cross-searching facilities. In addition, set up a general European archive for researchers with access to a subject-based or institutional archive.”

Accordingly, a long-term vision is to build a Europe wide Digital Repository infrastructure, which follows the principle of linking users to knowledge. This will form an integral part of the e-infrastructure for research in the future. Digital Repositories (DRs) contain today the full spectrum of scholarly materials, from theses, technical reports and working papers to digitised text and image collections and they can contain sets of primary research data. Institutional Repositories (IRs) present a specific form of a Digital Repository, where a significant fraction of the content is created by the research institution such as universities, research institutes, national laboratories, etc. The future Europe wide Digital

⁶ <http://www.jstor.org>

Repository Infrastructure can thus be thought of as a virtual network of physically distributed and locally maintained repositories from all countries in Europe.

This pan-European knowledge infrastructure should be planned to be built on the existing local, national and trans-national knowledge infrastructures and grid software and infrastructure. Its design should avoid any sharply defined boundaries. It should be easily federated with other knowledge infrastructures in other parts of the world, and it should be easy to open its benefits to other potential user areas such as e-health, e-learning, e-government and others.

The new knowledge infrastructure should be open to all potential service providers, both from the academic community and from the commercial sector. This will lead to greater creativity and richness of services and facilitate new forms of public/private partnerships in the knowledge service area.

The plans for this infrastructure should take into account the need for sustainability. There should be close working with other key players: with the grids community, to identify common middleware, to follow common software quality assurance processes, and also to make use of their experience in re-use of software; with the open source community since much of the software in use has been provided by them; with the library community since preservation has a key role in sustainability.

In this scenario, DILIGENT can act as the digital library infrastructure that facilitates knowledge sharing and remote co-operation in e-Science.

3.3 Outcomes

This chapter illuminates the functional perspective of Digital Libraries and Scientific Repositories by extending the analysis of D4.3.1 and introducing the scientific repository perspective:

- Section 3.1 reports on trends in DL and Scientific Repository domains, and focusing on a functional comparison of three well known existing solutions from both areas; DSpace, Fedora, and aDORe for repository applications and OpenDLib, OSIRIS/ISIS, and DAFFODIL for DL systems. The comparison of target users, information space, overall functionality, quality of service, and underlying architecture has been carried out with respect to the DELOS Reference Model for Digital Libraries [8].
- Section 3.2 focuses on the scientific repository perspective of recent studies which try to find a model for scientific publication that could involve readers, authors, publishers, and funding bodies. In order to effectively link users to knowledge, the short-term but still sustainable public access to publicly-funded research results by the support of open archives is considered crucial. In this vision of the creation of Europe-wide Digital Repository infrastructure, which will form an integral part of the e-infrastructure for research in the future, DILIGENT can play an important role as it can act as the digital library infrastructure that facilitates knowledge sharing and remote co-operation in e-Science.

4 IMPACT ON COMMUNITIES

This chapter introduces the end-users communities' point of view in the analysis of the future market of DILIGENT service/product.

Two user communities and one major content provider are already involved in this project by means of relevant representatives:

- 1) ESA (European Space Agency) as leader of Earth Observation Science Community
- 2) SNS (Scuola Normale Superiore di Pisa) as representative of Humanities community, including High Education Institutions
- 3) RAI (Radio Televisione Italiana) as representative of content provider and stakeholders of public content

An initial analysis of impact on such communities was made by providing a short set of questions such representatives.

In the next version of this document, D4.3.3b, foreseen by M31, the impact analysis will be expanded including other members of these communities.

4.1 Earth Observation Science: the ESA case

The European Space Agency (ESA) is nowadays facing new requirements in terms of content and applications management that are often quite different in target and scope, ranging from Earth Observation (EO) to space exploration and astronomy sectors.

In this framework, traditional digital libraries (DLs) are not suitable anymore to satisfy emerging activities like, for example, assessments and planning responses to environmental accidents. To face such situations DLs must provide further facilities for storing, managing and accessing multi-type information, for making community specific applications and high computing-storage capabilities available, and for responding to proper on-demand aggregation and interoperation of data and services within user-defined processes.

That's why basically ESA is part of the DILIGENT project where part of the test-bed should serve a typical application scenario of EO, named ImpECT (Implementation of Environmental Conventions). In particular the DLs that ImpECT users will be able to create on-demand should support in terms of data and services the building of technical and periodical environmental reports, mostly related to international and regional conventions on marine pollution.

4.1.1 Community characterization

The ImpECT community is formally constituted by the 21 states signatories to the Barcelona Convention (an agreement dedicated to protecting the Mediterranean marine and coastal environment) along with other actors like coastguard offices, the United Nations Educational, Scientific and Cultural Organisation's Intergovernmental Oceanographic Commission (UNESCO IOC) and the Regional Marine Pollution Emergency Response Centre for the Mediterranean Sea (REMPEC). Other users are non-governmental organisations active in the Mediterranean including the World Wildlife Fund (WWF), the International Tanker Owners Pollution Federation Limited (ITOPF), and the Mediterranean Oil Industry Group (MOIG).

4.1.2 Best practices

Currently there are no practices in the semi-automated management of complex environmental reports (namely, *living documents*): that is a typical EO activity still played manually by EO experts. What nowadays is available concerning EO resources are a lot of dispersed and heterogeneous information sources and services often referring to different representation standards (see e.g. <http://www.opengeospatial.org/>, <http://www.iso.org/>, <http://dublincore.org/>): such a situation leads to a very time-consuming activity in retrieving and collecting pertinent and useful information and in maintaining over time reports that must obviously evolve with facts.

Environmental reports are nowadays created by manually aggregating and merging existing information, usually coming from different heterogeneous sources that have to be properly discovered and uniformly accessed. Moreover the collected information has often to be *coherently* integrated with other *pertinent* information; part of the information to manage may also be generated on-demand through procedures that often need to access and process huge amount of data.

Given the relevant social role of these environmental reports, once published, they have to be constantly maintained up to date. Maintaining this kind of reports may concern the re-generation of summary maps used to show environmental features and characteristics distribution, such as chlorophyll or vegetation indexes: those maps may then need to be re-built at each report versioning. Re-building of maps may involve a lot of expertise from the human-side and also a lot of facilities from the underlying document management system that are not there yet, like instantiating and running ad-hoc procedures and algorithms dealing with whatever amount of data locally available or either dispersed among different sites.

To show some relevant examples of current practices in collecting EO material and content provisioning initiatives the NASA CEOS IDN (<http://idn.ceos.org>), the European Environment Agency (<http://www.eea.eu.int/>), Medspiration (<http://www.medspiration.org/products/>) and the FAO GeoNetwork (<http://www.fao.org/geonetwork/>) can be cited. Moreover the following ones are all placed and maintained at the ESA's European Space Research Institute (ESRIN):

- EO ESA web portal (<http://www.eoportal.org>). This is a collection of heterogeneous information objects collected from a large number of providers. For instance, it contains classic documents like research studies and meteorological papers, EO products provided by ESA, DLR, NASA, and others, images like the satellite image of the Black Sea showing swirling blooms of phytoplankton colouring the surface waters blue and green, various maps presenting data on different studies like oil pollution, burn scars, cloud cover, etc.
- EO Grid on demand system (<http://eogrid.esrin.esa.int>). Examples of EO products that Grid on demand is able to generate are the Chlorophyll-1 measure corresponding to the algal pigment index 1 expressed as a chlorophyll concentration in mg.m⁻³, the mosaic made up of true colour images using four out of 15 MERIS spectral bands taken from Envisat with data combined from the selected separate orbital segments with the intention of minimising cloud cover as much as possible by using the corresponding data flags, and vegetation indexes measuring the amount and vigour of vegetation at the surface.

4.1.3 The expected impact of DILIGENT

Within an EO application scenario, the facilities such as those envisaged by a DILIGENT user interface, as described in the project user requirements document and functional

specifications, should be able to address all the following key factors, of relevance to the ImpECT community:

- management of very large and distributed virtual organisations;
- seamless access to and handling of distributed and heterogeneous data and services;
- on-demand and efficient processing of huge amounts of information;
- support the definition of ad-hoc user defined workflows of services together with scalable and reliable executions;
- storage of the derived data as well as of the dependencies between them with support to knowledge preservation;
- traceability of the operations performed.

To refer to a concrete plus in gaining DILIGENT, EO experts should be able to use facilities to maintain reports for building ad-hoc compound services to generate, manipulate and modify data. Such facilities would be in charge of managing the available resources, both storage and computational ones, to allow their usage as effectively and efficiently as possible. Workflows may involve any of the EO services, e.g. vegetation index or mosaics generators, web map and web features services, data analysis and visualisation services, numerical models services, hazards planning services. Such EO and general services separately or aggregated and chained within automated workflows will allow to interactively maintain environmental reports, realising the notion of *living documents*.

Once DILIGENT will be able to offer a complete set of services fulfilling that list of requirements the impact on the EO content management and exploitation will be of enormous importance, both due to technical innovations and social impact. What ImpECT users are mostly waiting from DILIGENT is the concrete ability to create on-demand ad hoc-user defined DLs, containing specific data and services, where to subsequently store and maintain also their ad-hoc defined environmental reports.

4.2 Humanities: the SNS experience

The *Scuola Normale Superiore* (SNS) of Pisa was established under a Napoleonic decree on 18 October 1810, as a branch of the École Normale Supérieure of Paris for Italian language territories, but its activity only began in 1813. Presently, the SNS is inserted within the Italian university system with a special set of regulations. It has as its main aim top-level university training and research activities. It is distinguished by the qualitative development of its university instruction and is characterized by the rigorous application of standard principles. The presence of three Nobel prize-winners among the slightly fewer than 5000 students of the SNS since its foundation, confirms the high level of this institution not only in the Italian, but also in the international academic context.

The *Centre for Literary Tradition Texts and Images Electronic Processing* (CTL) is one of the research center of the SNS. Founded in 2000 and directed by Prof. Lina Bolzoni, it aims at realizing research projects, which, by making use of computer science technology, investigate the complex relationships between words and images present in the literary tradition. The CTL has a top level scientific Committee and collaborates with scholars from Italian and foreign universities and cultural institutions, among them the École Normale Supérieure, the Collège de France, the New York University and the University of California-Los Angeles. The centre co-ordinates research projects aimed at rediscovering the large areas of interaction between word and image that have characterized the literary tradition. The creation of electronic archives of words and images becomes from this viewpoint also an instrument that allows the reconstruction of various forms of

interweaving between the linguistic and the figurative codes, as well as to analyze the less known areas of inter-expressive experimentation. It is engaged in numerous national and international research projects.

4.2.1 Community characterization

The community consists of researchers of a research project, named ARTE (in Italian: Applicazione di Ricerche e Tecnologie di Editoria digitale) and funded by MIUR, the Italian Ministry for Higher Education and Research. This is a community of scholars, distributed all over the world (almost 20 professors and researchers) who have decided to start working together in order to set up the basis for a new research discipline that merges together experiences from the medical, humanity, social science, and communication research areas. This community is very interested in the Digital Library technologies, because they consider a DL as a medium for communication and cooperation besides seeing it as an instrument to create a common knowledge base. The ARTE community is represented in DILIGENT by the CTL-SNS. Other institutions of the ARTE community are:

- Universitat de Barcelona, Proyecto Boscàn (Catálogo de las Traducciones Españolas de Obras Italianas hasta 1939) - <http://www.ub.es/boscan/>
- University of Utrecht, Emblem Project Utrecht - <http://emblems.let.uu.nl/emblems/html/>
- Universidade da Coruña, Research Team on Hispanic Emblematic Literature - <http://rosalia.dc.fi.udc.es/emblematica>
- University of Glasgow, HATII (Humanities Advanced Technology and Information Institute) - <http://www.hatii.arts.gla.ac.uk>
- RAI Medita, Italian Digital Multimedia Library from RAI archives - <http://www.medita.rai.it/index.htm>
- Studio Azzurro Produzioni - <http://www.studioazzurro.com/>
- Warburg Institute (<http://warburg.sas.ac.uk/>)
- Biblioteca Digitale Italiana (<http://www.internetculturale.sbn.it/genera.jsp>)
- Stanford University Library (<http://library.stanford.edu/>)

4.2.2 Best practices

The CTL has already experienced the use of a DL for their research activities. In fact, in the recent past a DL managed by the OpenDLib 1 system has been set up as a result of collaboration between CTL and ISTI-CNR (see also 3.1.2). The CTL DL manages by OpenDLib two very different collections, namely the "Atlas of Memory Images" collection and "Narrated Dreams in Modern Literature" collection. The former offers the images contained in treatises on the art of memory of the XVI Century; the latter contains literary texts that narrate or describe dreams (the related digitalized images are currently under preparation).

CTL researchers use habitually several Digital Archives such as:

- Gallica (<http://gallica.bnf.fr/>)
- Biblioteca Emblematica (<http://dinamico.unibg.it/cav/emblematica/>)
- Arts and Humanities Data Service – AHDS (<http://ahds.ac.uk/>),
- Biblioteca Digitale Italiana (<http://www.internetculturale.sbn.it/genera.jsp>)
- Warburg Institute (<http://warburg.sas.ac.uk/>)
- Stanford University Library (<http://library.stanford.edu/>)

4.2.3 The expected impact of DILIGENT

Scholars engaged in the ARTE project are presently using their digital library for setting up digital collections and making research on texts and images. They consider their experience very successful, although it is limited by the fact that the ARTE project doesn't possess computational and storage resources for handling large sets of multimedia documents. For this, they are very much attracted by the possibility of exploiting DILIGENT as a means to provide themselves a cost-effective instrument for setting up multimedia knowledge repositories equipped with a number of services specifically tailored to their needs. In particular, they are attracted by the possibility of creating a digital library on demand, i.e. a digital library for a temporary use - for example for organizing the material to be used by course students or to be exhibited/discussed in a workshop. This means that the creation of a DL, or a DL collection, should be made by aggregating content and services already available on the DILIGENT infrastructure, in order to reduce the costs and the time needed to create the environment capable to satisfy the specific needs of the ARTE project.

In order to achieve their objectives the ARTE researchers need to establish a common background knowledge base. DILIGENT will be experimented as a means to provide them, in a short time frame, a cost-effective instrument for setting up VDLs, i.e. common multimedia knowledge repositories equipped with a number of services, specifically tailored to the needs of this community. These VDLs will enable the ARTE researchers to exploit the benefits of a DL despite the limited duration and funding resources available for the ARTE project. DILIGENT will also be used experimentally as a means for supporting another typical activity of the ARTE project members: the organization of courses. Specific VDLs that address the knowledge needs of the students can be created by re-using material maintained in the registered archive resources of the DILIGENT infrastructure. These VDLs automatically update their content following the changes in the original archives. As a result, the students of each course have access to the most updated material on the topic of each course. In addition to the capability of creating VDLs, the DILIGENT flexible global infrastructure will provide to the ARTE team an advanced, content-based multimedia information retrieval searching functionality across heterogeneous and geographical dispersed collections. Moreover, through the underlying Grid infrastructure, DILIGENT will permit the exploitation of sophisticated processing functions for analyzing and comparing multimedia data, allowing the discovery of hidden structures or semantic relationships that reveal important aspects of the association between images and words, both at content creation and query time.

4.3 Content Provider: the RAI point of view

RAI Radiotelevisione Italiana is Italy's national broadcaster with three terrestrial channel, a variety of dedicated satellite channels, digital terrestrial channels and audiovisual VOD broadband services.

In Diligent, RAI is participating through RAI Educational, the department providing educational programs and services to specific user communities such as the Italian schools, and to the interested public. Besides program slots on the general channels, content delivery is mainly effectuated through two digital satellite channels RAI Edu 1 and RAI Edu 2, and to numerous streaming video services provided through more than 20 websites accessible within the portal www.educational.rai.it. A specific videoservice for schools is offered through the audiovisual Medialibrary MEDITA at www.medita.rai.it, with more than 2500 pedagogically relevant videos, offered through a database searchable by keyword, by subjects, by age groups and school types.

4.3.1 Community characterization

MEDITA is targeting the whole of the Italian school system (more than 7000 schools) at the moment through its broadcast on the satellite channel, and a server-client based streaming web site containing all approx. 2500 videos. Schools/teachers have to register on the website and get then full-access to streamed videos as they like and to upload eventually material related to a specific video-topic which will then be available to other users.

4.3.2 Best practices

Generally speaking, due to the technological achievements (iTV, IPTV, P2P, Video-Podcast) the audiovisual sector is undergoing currently what could be called a complete change of the paradigm of content distribution: whereas broadcasters used to offer centralized, linear, time-dependent distribution, the future will bring user-driven on-demand services. This is sometimes referred to also as "the death of tv as we know it" (cf. an important study done by IBM Global Services www.ibm.com/bcs/endoftv). For broadcasters like RAI, there is a real challenge to master this swift from the old to the new paradigm without losing to much of their traditional market potential, and of course keeping alive the mission and ideals of a public interest broadcaster (Best practices in RAI are probably www.medita.rai.it; and RAI's VOD-service www.raidlick.rai.it).

A much discussed case on a European/global level which might become "best practice" soon is a BBC pilot (2005) using P2P technology for video delivery to consumers. A fundamental reason for the BBC to explore the potential of P2P lays in the relatively low distribution costs, when it comes to high user numbers. For more information on this case, cf. <http://informatv.com/opinion/2005/11/impfirst> and <http://www.bbc.co.uk/imp>.

4.3.3 The expected impact of DILIGENT

As a player in the audiovisual industry, RAI's expectation regarding DILIGENT's output obviously focused initially mainly on possible progresses in the management of audiovisual data (at least as a major component in a multi-data and media library environment) through the use of grid-technology. One concrete "vision" - which RAI expressed in a presentation at the Arte Kick-off Meeting (<http://dlib.sns.it/bscw/bscw.cgi/0/25281>) - regards the possibility to substitute sometime in the future the actual cost-intensive server-client architecture of MEDITA, RAI's video library for schools, with a grid based interactive and multi-centered distribution network, where Italian schools form an educational "Virtual Organisation" sharing a multimedia DL managed as a grid, in the sense in which this concept has been defined in the DILIGENT functional specification.

Despite the achievement of the above objective through the DILIGENT infrastructure is feasible from the theoretical point of view, its practical realization requires more investigation and the implementation of additional, more specific, services for video handling on the grid.

4.4 Outcomes

This chapter introduces the point of view of the end-user communities in the analysis of the future market of DILIGENT service/product. An initial exemplary analysis of best practices and foreseen impact has been carried out on the representatives and content providers of the two user communities which are involved as partners in the project:

- ESA (European Space Agency) as leader of Earth Observation Science Community.
- SNS (Scuola Normale Superiore di Pisa) as representative of the Humanities community, including High Education Institutions,...

- RAI (Radio Televisione Italiana) as representative of content providers and stakeholders of public content.

The analysis of impact on such communities was made by providing a short set of questions to the representatives of communities. Most worth noting are the expected impacts of DILIGENT on

- The Earth Observation community (represented by ESA) which anticipate that DILIGENT will have enormous effect on their content management and exploitation capabilities, leading to technical innovation and social impact. This will be achieved by DILIGENT enabling the EO community to manage large and distributed VOs, integration with existing EO services, and facilitating the creation and maintenance of environmental reports which will become powerful *living documents* to be executed on.
- The Humanities community (represented by SNS) which understands DILIGENT as a cost and time effective instrument for the on-demand creation of digital libraries and multimedia knowledge repositories which provides a platform for collaboration of scholars and students in courses and workshop. Due to the embracement of grid technology, DILIGENT is anticipated to supply advanced storage and computation resources for large sets of heterogeneous multimedia objects – functionality which is missing in state-of-the-art DL systems presently used by SNS.
- As a major player in the audiovisual industry, RAI, expects DILIGENT results to help broadcasters to master the upcoming challenge of the paradigm shift in content distribution. As one example it is expected that DILIGENT will enable RAI to cost-effectively manage and distribute educational content to collaborating schools.

In the next version of this document, D4.3.3b, foreseen by M31, the impact analysis will be expanded including other members of these communities.

5 GRID AS ENABLING TECHNOLOGY

This Chapter reports on the monitoring of grid related markets and technologies as the key enablers for the development and later commercial take-up of DILIGENT.

Grid technology is the main enabling technology for DILIGENT and so a deep constant analysis of the grid market trends (section 5.1) and of the evolutions of other grid-middleware and standards (section 5.2) is done.

In appendices 9.1-9.4 we report on the status and activities of grid-related standard bodies and the description of EGEE II, the project successor of EGEE, which will take the responsibility of the maintenance and evolution of gLite middleware.

5.1 Overview of the market and trends

The Grid is normally seen as a framework for "flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources" [14]. It allows researchers in different administrative domains to use multiple resources for problem solving and provides an infrastructure for developing larger and more complex applications potentially faster than with existing systems.

Grid technology is being adopted by the scientific research community, and by certain vertical commercial market segments. The most notable vertical commercial sectors are banking/finance and biotech/pharmaceutical. In contrast to research, these deployments use professionally supported middleware from independent software vendors. There is a growing interest in Grid and related technologies of virtualization, automation, Service-Oriented Architectures and utility services. Adoption in commercial settings is in early adopter high-performance computing segments – commercial users of mainstream IT are slower to adopt. In general, commercial users tend to use supported Grid middleware products, and academic users tend to use open source tools. In general terms, the technology market has moved from a initial period of low awareness of the Grid concepts to a second phase of basic understanding of what the Grid is and what it can provide and finally to a third period where organisations are able to distinguish the different types of Grids. On the other hand, in more technical terms, the Grid has also evolved over this period. Here the evolution was from local infrastructures (running large applications on high-performance clusters) to dynamic virtual environments supporting heterogeneous requests from different user communities. This evolution, summarised here below, is well documented in a series of reports produced by Quocirca (<http://www.quocirca.com>), a Business and IT company which is investing effort in studying the Grid evolution in the recent years.

At a certain point of this "maturation" process, the real potential of the Grid started to be understood and became a goal for many groups. Most of these groups, share a common vision – that moving towards Grid is inevitable. Although many problems still exist, they are not seen as a blocker to the decision to move for a Grid architecture. The speed of this "maturation" process is also surprising. In early 2003, Quocirca believed that mainstream Grid computing was still around 5 years away. In 2005, 20% of companies were already utilising or experimenting with Grid, pointing to mainstream adoption within the next 2 years.

This rapid take-up of Grid has been helped by the speed of adoption of open standards by all the major players in the market. However, this strength may also end up being Grid's weakness. We are already seeing the early Grid standards being upgraded, and we also have multiple bodies defining slightly different versions within the market. If users cannot depend on a fully standardised, interoperable environment, we may yet fall back into the problems of yesteryear, where de facto standards (those already in use within the end user environment) were modified by de jure groups (standards committees, often driven

by different agendas), and multiple solutions led in divergent directions to solutions that ended in evolutionary cul-de-sacs. For example, apart from the core Web services components (SOAP, WSDL, etc) all the other existing standards are still not seen as de facto standards. It is therefore extremely important that vendors and organisations keep an eye on standards evolution within the Grid markets and make an effort in achieving standardization.

Apart from standards, and focusing in the market itself, Grid is being seen more often as part of the natural evolution and as the leading-edge that will bring market differentiation through its added value. Quocirca has studied how Grid computing is being adopted by companies and what knowledge these companies have on grid solutions. Figure 5.1 shows this evolution.

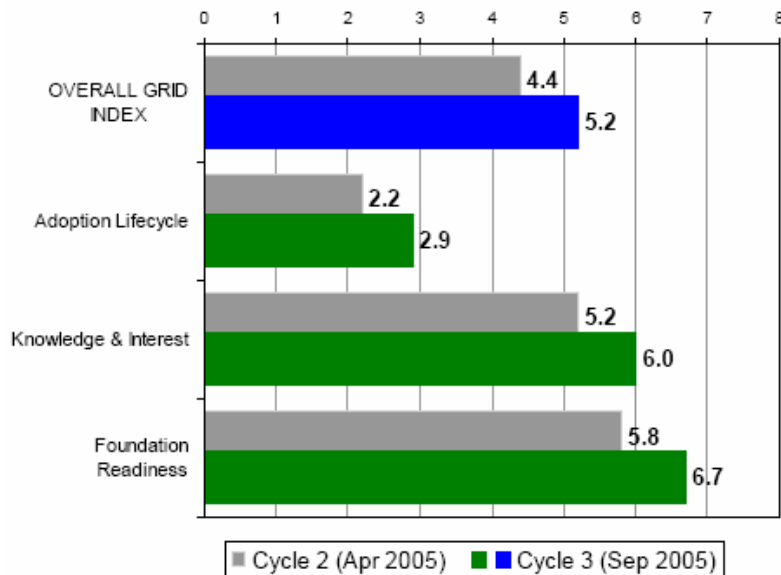


Figure 5.1: Grid uptake in IT companies (<http://www.quocirca.com>)

Here we see reasonable growth across the board. Foundation readiness continues to grow, and is now at a level where the majority of companies have a capability to begin looking towards Grid implementations. Knowledge and interest have also grown, and this has led to a readiness to look at different forms of Grid computing. The adoption index has also grown – but is still relatively low. This is not surprising, as other findings in the research point to companies still not regarding Grid as a mature solution, of standardisation still having to be tied down, of possible problems in introducing Grid into already complex environments – and also around how such a concept should be budgeted for, funded and owned within a highly dynamic organisational structure.

Essentially, Grids are now a “public” concept. The coordinated affirmation of open standards, web services and service oriented architectures will “deliver” a highly dynamic and efficient infrastructural architecture. With these foundations ready and with the general Grid awareness, organisations are gradually moving towards Grids.

5.2 gLite vs. other solutions

A large number of Grid research projects have started in Europe in the last decade. Their common objective is to virtualise resources empowering individuals and organisations to

create, provide access to and use a variety of services, anywhere, anytime, in a transparent and cost-effective way.

DILIGENT adopted gLite, supported by the EGEE project, as the grid middleware solution most suitable for its purposes.

There was no choice between research projects results and commercial solutions since the presence of commercial European entities in the Grid product market was almost non-existent: all Grid computing middleware were the result of research projects based on the North-American Globus Toolkit.

GLite was chosen as the grid middleware solution for DILIGENT since it is a complete solution providing a full set of grid functionality needed to satisfy the DILIGENT requirements. Even if still evolving - and open to support DL requirements - the gLite middleware runs today in more than 200 sites linked to the EGEE infrastructure. It is extensively tested and largely adopted, therefore stable and more likely to be maintained in the long term compared to other solutions. Through gLite, DILIGENT has potentially access to a large number of shared resources. It has to be highlighted that the EGEE project also works alongside several other major European projects, collaborates with Industry and is in the process of defining a long term sustainability plan for its infrastructure and middleware from which DILIGENT will also benefit. See appendix 9.1 for more details.

5.2.1 Other Grid Middleware Initiatives

This section aims at identifying the major and current solutions in terms of what is called *Grid middleware*. Grid middleware refers to the security, resources management, data access, instrumentation, policy, accounting, and other services required for applications, users, and resource providers to operate effectively in a Grid environment: the middleware is meant to act as a sort of 'glue' to binds these services together.

A large number of Grid research projects have started in Europe in the last decade. Their major objective is to virtualise resources empowering individuals and organisations to create, provide access to and use a variety of services, anywhere, anytime, in a transparent and cost-effective way. This will realise the vision of a knowledge-based and ubiquitous utility. Anyway the presence of commercial European entities in the Grid product market is almost non-existent: all Grid computing middleware are the results of research projects based on the North-American Globus Toolkit.

EGEE represents the larger collaboration bridging together many distributed resources to build an extended and powerful infrastructure. EGEE integrates several best practices from other EU research projects and provide an open-source Grid middleware, *gLite*. The EGEE infrastructure builds on the EU Research Network GEANT and exploits Grid expertise that has been generated by projects such as the EU DataGrid project, other EU supported Grid projects and the national Grid initiatives such as UK e-Science, INFN Grid, NorduGrid and US Trillium. The infrastructure aims at providing interoperability with other Grids around the globe, including the US and Asia, contributing to efforts to establish a worldwide Grid infrastructure.

Here below follows a list of the major European and non-European commercial and non commercial initiatives on Grid middleware services: in particular the table summarizes the set of solutions investigated, specifying for each of them the web link of reference and to which extent it can effectively be considered relevant to Diligent, i.e. if it can provide all the core grid functionalities needed by Diligent, *full* solution, or only part of them, *partial* solution. More specifically the coverage of the grid functionalities is meant to be *full* within a Diligent context when *security services*, *information and monitoring services*, *job management services* and *data services* are there in the specific solution. Examples of

relevant partial solutions to Diligent are GridFTP, GridICE and OGSA-DAI, described in Section 9.2 of this document and in its previous version (D4.3.1).

In case a solution has already been cited within the previous version of this deliverable (see column 'Referenced in D4.3.1') it will not be further described while all the *new* solutions are addressed in Section 9.2.

More information and references on the standards adopted by those solutions can be found in Section 5.2.2.

Solution Name	Web page	Description	Coverage
Condor	http://www.cs.wisc.edu/condor/	D4.3.1	partial
Deisa	http://www.deisa.org/grid/initiative.php	new	partial
EnFuzion	http://www.axceleon.com/solutions.html	new	partial
GFarm	http://datafarm.apgrid.org/	new	partial
gLite	http://glite.web.cern.ch/glite/	D4.3.1	full
Globus Toolkit	http://www.globus.org/	D4.3.1	full
GPT	http://www.globus.org/grid_software/packaging/gpt.php	new	partial
GridAssist	http://dsgrid.xs4all.nl:8080/gridassist/index.html	new	partial
Gridblocks	http://gridblocks.sourceforge.net/	new	partial
GridFTP	http://www.globus.org/grid_software/data/gridftp.php	new	partial
GridICE	http://infnforge.cnaf.infn.it/gridice/	new	partial
Grid MP	http://en.wikipedia.org/wiki/Grid_MP	new	partial
GSI plugin for gSOAP	http://www.gridlab.org/WorkPackages/wp-5/guide/gsoap.html	new	partial
MS .NET Grid	http://www.nesc.ac.uk/action/projects/project_action.cfm?title=145	new	partial
OMII	http://www.omii.ac.uk/	new	partial
NMI	http://www.nsf-middleware.org/about.aspx	new	full
OGSA-DAI	http://www.ogsadai.org.uk/	D4.3.1	partial
OSG	http://www.opensciencegrid.org/	new	partial
pyGMA	http://www.didc.lbl.gov/pyGMA/pyGMA_top.html	new	partial
TUPELO	http://dlt.ncsa.uiuc.edu/wiki/index.php/Tupelo_2	new	partial
UNICORE	http://www.unicore.org/	D4.3.1	partial
XGrid	http://www.apple.com/macosx/features/xgrid/	new	partial
iGrid	www.crosswalkinc.com	D4.3.1	partial
NAREGI	http://www.naregi.org/index_e.html	D4.3.1	full
VDT	http://vdt.cs.wisc.edu/	new	partial

Table 3 - List of other Grid middleware initiatives

Diligent has adopted gLite as its grid middleware solution to build on top. gLite is a full solution that provides all the grid basic functionalities needed by Diligent. In addition gLite exploits experience and existing components from other solutions mentioned above such as Condor, GridFTP and Globus Toolkit. All these solutions have been formally described within the previous deliverable version (D4.3.1) and in the current document (see Sec 9.2).

5.2.2 Grid Standards and their evolution

The recent adoption of Web services, while bringing significant benefits, has also produced a heterogeneous environment for application developers. Grid services developers always prefer to conform to widely adopted conventions and standards. In addition, the Grid vision requires protocols that are not only open and general-purpose but also standard.

The solution is therefore the standardization. It is standardization that allows potential collaborators to establish resource sharing arrangements quickly and easily with any interested party. It is standardization that allows organizations to establish resource sharing contracts routinely for acquiring resources on demand, thus avoiding the need to build expensive data centres designed to handle peak loads but that remain underutilized most of the time. Standard solutions can allow us to move away from today's often incompatible and non-interoperable distributed systems, toward a model where computing and data capabilities are available as standardized, interchangeable commodities.

This standardization effort that is leading the future of Grid computing is being carried out by several groups, consortiums, communities and forums. Of all consortiums currently working in Grid standards, some are already established groups now adapting their efforts in this new direction, while others are recently formed groups, fully Grid oriented. The most relevant and well known are described in appendix 9.3.

Another aspect which is taken into consideration is the standard followed by each solution. The more important ones, either still emerging or widely used, are presented in the following table. Some of these standards were already described in the previous deliverable [D4.3.1 "Market and Technology Trends Analysis"](#) where an initial overview of the grid market was given. In this deliverable, new standards are described while other standards already described in D4.3.1 are updated. The description of these new and updated standards can be found in the Appendix 9.3.

S. Body	Standard	Ver.	Description
DMTF	Web Services Management (WS-Management)	1.0	New
GGF	Open Grid Services Architecture (OGSA)	1.0	D4.3.1
IETF	Grid Security Infrastructure (GSI)	-	New
	Lightweight Directory Access Protocol (LDAP)	3.0	New
	Secure Socket Layer (SSL)	3.0	New
OASIS	Web Services Security (WS-Security)	1.1	D4.3.1
	Web Services Resource Framework (WSRF)	1.2	Updated
	Web Services Notification (WS-N)	-	Updated
	Universal Description, Discovery, Integration (UDDI)	3.0	New
	Web Services Reliability (WS-Reliability)	1.1	New
	Web Services for Remote Portlets (WSRP)	1.0	New
W3C	Extensible Markup Language (XML)	1.1	New

	Simple Object Access Protocol (SOAP)	1.2	D4.3.1
	Web Services Description Language (WSDL)	1.1	D4.3.1
	Web Services Addressing (WS-Addressing)	1.0	D4.3.1
WS-I	Web Services Interoperability Basic Profile (WSI BP)	1.1	New

Table 4: Standards in grid computing

The standards presented in this table were selected mainly due to the considerable consensus they gather among different grid computing domains. It is also clear that the “acceptance level” is not the same for all selected standards. Some of them – such as XML, SOAP, etc – are already in a mature stage and are unconditionally used. Other – such as WSRF, WSRP, etc – are still as emerging. However all of them are seen as the correct standards that are and will lead the grid computing developments. DILIGENT is therefore adopting some of these standards and monitoring the progresses done by the standardization bodies.

5.3 Outcomes

This chapter has analysed the status and evolution of grid technology around the world. In fact the relevance of this technology has no frontiers and the competition (and collaboration) is out of the European frontiers.

There are some interesting findings that need to be highlighted:

- the start of EGEE II project, that will give to DILIGENT the assurance of at least another two years of a grid-middleware fully supported and a strict collaboration in adequate it on DILIGENT requirements;
- EGEE sustainability plan and the foreseen implementation of EGI;
- the lack of any commercial solution that could compete with the gLite adoption in terms of dependability and support;
- The convergence of grid-related standards towards a web-service based notion of grid service, already adopted by DILIGENT during its early design phase;
- The merge of Global Grid Forum (GGF) and Enterprise Grid Alliance in a unique entity (Open Grid Forum, OGF) that will afford the challenge to realize common standards for all aspects of grid technology. As for GGF, OGF will be one of the major communities in which DILIGENT will discuss technical choices and submit requirements.

6 EMERGING SCENARIOS IN DIGITAL CONTENT MANAGEMENT

This chapter identifies and analyses the reference market for DILIGENT services/products: the Digital Content market.

DILIGENT is focused on research communities, thus aiming at better supporting existing communities' needs in performing their research activities.

Nevertheless this context is influenced by a great emerging scenario in the ITC world: the convergence of content, network and computing infrastructures, as in V. Reding speeches related to i2010 initiative.

"i2010 identifies digital convergence as the main driver of change and aims at ensuring that the EU will fully benefit from the opportunities and prospects for strengthening the Single Market. i2010 sees four challenges for convergence: speed, rich and diverse multilingual content, interoperability and security".

[reference:

http://europa.eu.int/information_society/eeurope/i2010/docs/annual_report/com_2006_2_15_en.pdf]

Such convergence will have an impact not only on the commercial field but even in the research context, where new needs may arise and new ways of doing research would be possible.

Such convergence needs to be taken into account when we define the sustainability strategy for DILIGENT that will possibly happen starting from 2008.

Currently some issues of digital convergence are already a reality. Investments in networks resumed in 2004 and 2005 and broadband subscriptions grew by 60% in 2005. Telecom and cable operators are offering converged services, such as 'triple-play' services⁷ or TV-over-IP. But revenues from these services do not compensate for the loss of voice revenues, and overall growth in the revenues of electronic communications is slowing down. Even on the radio market, for many years an "analogical fort", the first digital radios are appearing..

The objective of this chapter is to have a clear understanding of the state-of-the-art of Digital Contents market, such emerging scenarios, possibly new value chains and sustainability best practice in the light of digital convergence.

6.1 Pervasive Media: the Content market in 2010

In the Framework Program (5th 1999-2002) one of the main fields for investment was the format transition, from analogical to digital. Great efforts were made to allow digital handling of analogical or physical content. Digitisation was a great deal to reduce the cost of management contents and of accessing them by almost any person in the European Community (and the rest of the world).

The actual Framework Program (6th 2003-2006) managed to put the convergence of content, networks, and services in reality. Some initial attention was made in the definition of simple methods to access the same content through several different media.

Such enabling environment, often referred to as AMI (Ambient Intelligence) will be one of the main deals in the next 5 or 6 years (about 2010) along with the creation of a unique pervasive infrastructure that will deal with all the aspects of managing any content.

⁷ Voice, data and video services offered by a single operator/service provider

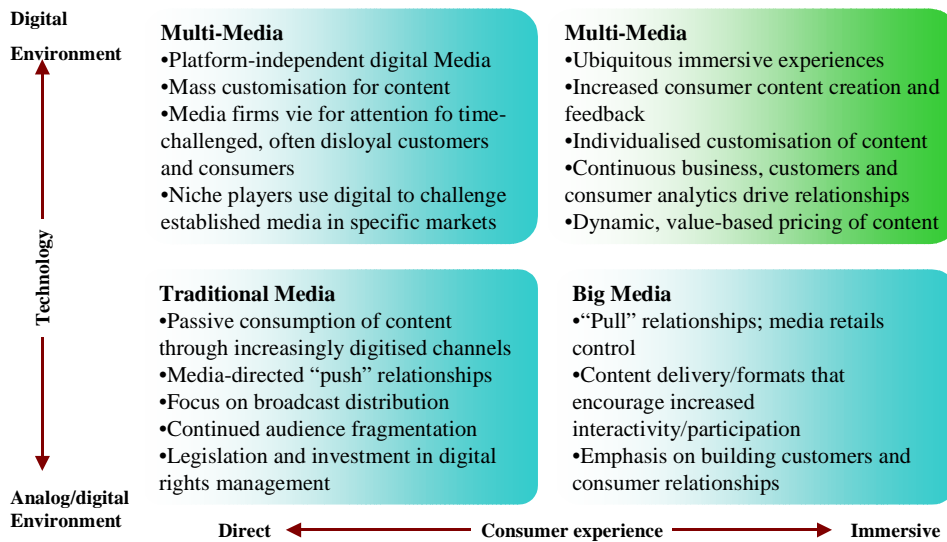
This is the idea of Pervasive Media: a three tier environment made by the human, the interface device (allowing the access to content), and the infrastructure (enabling “intelligent” supply of content).

The infrastructure is the layer with which DILIGENT is dealing with, and the digital convergence will enable it to support the functionalities required by the Pervasive Media.

This evolution is well described in a perspective picture from IBM Business Consultants that illustrates the growth of new behaviours in media experience and consumption to the next period.

In the vertical axis the growth of technology is represented, essentially from legacy analog systems to the increasingly digitised environment. Along this axis content and formats are moving from producer and distributors formats toward an open market for delivery outside the control of the industry (i.e. the MP3 case).

The horizontal axis shows the trends in the experience of content consumption, whose final end is the immersive environment of the future.



Source: IBM Business Consulting Services Analysis; IBM Institute for Business Value analysis, 2004

DILIGENT is well positioned on this evolution scale since it is based on the G2/EGEE II pair that supplies the main functionalities serving such digital convergence. From one point of view DILIGENT is complementing the effort made by such initiatives (i.e. G2 and EGEE II) to accomplish the management of any digital content.

6.2 What content for the Digital Content Market

Thus the main aspect that DILIGENT faces is the capability to manage the right content type.

The concept of Digital Content Market often refers to several categories of content types:

- 1) Entertainment, that includes:

- a. Mobile entertainment services such as: logos, backgrounds, screensavers, ring tones (monophonic, polyphonic, true tones); interactive services based on Text Messages (Short Message Service); community and dating on mobile phone; mobile gaming; other contents of ludic type distributed through mobile phone such as horoscopes, jokes, etc. Adults contents offered on mobile phone are included in this market;
 - b. Digital music; musical pieces in digital form, supplied both through internet and through mobile phone;
 - c. Infotainment segment, represented by the totality of digital contents connected to entertainment and, more generally, to the ludic areas, as well as the information connected to free time, distributed through the internet and – even if in lesser measure – through digital terrestrial television and IP-television.
- 2) Video, that includes the contents⁸ distributed on the following platforms:
- a. Digital terrestrial television;
 - b. Digital Satellite Television;
 - c. Internet (the offer of Pay-TV and video-on-demand distributed on this media);
 - d. Mobile phone⁹.
- 3) News, that includes solely general and financial news¹⁰.
- 4) Cultural Heritage segment that includes the investments connected to museum and digitisation of libraries¹¹, it is mainly related to digital version of physical artifacts whose digital copy need to be managed. This is the one of the two focus markets for DILIGENT.
- 5) Scientific Data and Information, this market is related both to scientific data and maps and information (paper, articles, textual information in general) often produced by international entities and distributed for free under certain conditions (this is the case of ESA and meteorological information). This is the other market that DILIGENT will primarily take into account.
- 6) Tourism, excluding the on-line reservation services, includes the so called infomobility services, that are information of a tourist nature or simply connected to the location of public services that can be supplied through a different number of media, including the mobile phone.
- 7) Education that is the component of school-book publishing distributed in digital format (CD-Roms or the so-called learning objects available through internet).

DILIGENT as a research testbed does not aim to cover all these markets, nevertheless most of this content type need to be considered to support the future research communities in performing their activities. In fact we cannot think to research communities as the actual involved in the project or in related initiatives (Physics or Biomedical community, meteorologists or Earth Science observation, Education or Researchers on emblems).

⁸ Including adult videos

⁹ Video on mobile is not included on "mobile entertainment" category to keep the focus on the content as opposite to the device.

¹⁰ Sport news, for their ludic nature are included in the Infotainment; while video news and trials are included in the video market

¹¹ excluding the profits connected to the on-line ticket sale, that are ascribable to the e-Commerce services

The sustainability of DILIGENT as a research infrastructure service will be in supporting any research communities even the one that will be created in the near future.

Having this in mind a first attempt to characterise the target market of DILIGENT, in terms of content type have been done. The results are collected in the Table below.

Market Content types	DILIGENT interest	DILIGENT coverage
Entertainment	N/A	
Mobile entertainment	NO	
Digital music	YES	Currently not taken into consideration
Infotainment	NO	
Video	YES	RAI is providing feedback on this issue
News	??	Potentially could be approached for financial and historical research activities
Cultural Heritage	YES	SNS main target
Scientific Data and Info	YES	ESA and Impect scenario
Tourism	NO	
Education	YES	Partially covered by the ARTE scenario

Table 5: Characterisation of target market

6.3 Openness vs. Rights protection

In the actual Digital Content Market two competing approaches for the creation and distribution of digital content are emerging.

One approach, advocated by conventional content companies from the movie and music industries, as well as by government officials, emphasised the need for digital rights management (DRM) technologies to "lock down" content.

They argued that DRM allows them to set specific limitations on the use of content, thereby facilitating commercial models such as individual downloads or full subscription services. DRM supporters assured delegates that their models were gaining marketplace acceptance as the content owners claimed that they stand ready to license their material for distribution on multiple platforms such as the internet and mobile phones. Moreover the IPR protection as resulting from using DRM and license schemas will allow authors to have resources for continue feeding their creativity.

The alternate approach focused mostly on user-generated content. Such content is the combination of easy-to-use technologies and widespread internet access that has unleashed an unprecedented amount of new creativity.

Rather than reducing their reliance on digital rights management, content owners blamed electronics makers for a lack of interoperability. These included discussions on Creative Commons, an alternate licensing system that has enabled individuals to make more than 60 million works available to the internet community; Flickr, the photo-sharing site that

features millions of digital photographs; and the BBC Creative Archive, which allows UK residents to re-use original content from that country's public broadcaster.

The popularity of weblogs, or blogs, also attracted considerable interest. David Sifry, the founder and CEO of Technorati, a blog search engine, reported that his company now tracks more than 27 million blogs, with 75,000 new blogs created every day. Moreover, the popularity of blogs is not strictly a North American phenomenon. Presently, Technorati tracked more blog postings written in Japanese than in English.

Notwithstanding the support for the DRM and user-generated approaches, both face threats that could hamper their development.

For example, Digital rights restrict the numbers of times users can copy music, and DRM supporters seemed ready to acknowledge that the technology has created consumer concerns. Many stakeholders admitted that major problems have arisen from the lack of interoperability that has made it difficult for consumers to transfer lawfully acquired content from one device to another. For example, the Napster music subscription service has been hurt by the inability to transfer songs to an Apple iPod. This lack of interoperability will not be solved unless legislators force consumer electronic makers to act for a "content neutrality" of devices, aims at which DRM makers are already ready.

One threat comes from the limitations of DRM-enabled content, which can severely limit or prohibit legal modification of material. These concerns were raised on May 2006 at a UK parliamentary committee hearing on DRM by well-respected organisations such as the British Library.

User-generated content also faces the threat of the two-tiered internet, which has garnered increasing attention in recent months. This refers to internet service provider (ISP) plans to restrict subscriber access to software such as BitTorrent, which is widely used to distribute user-generated content such as independent films or open source programs. The two-tier internet could also hamper the growth of tools used to locate user-generated content, since ISPs such as BellSouth, Verizon, and Telus have all raised the prospect of charging websites and services for the right to deliver content to their subscribers.

All these concerns will affect heavily DILIGENT sustainability since the actual position of the project with respect of IPR is that of "don't care" due to the research nature of the project and contents managed. Nevertheless this will become soon as a reality when will be realised that content security and authorization and authentication policies alone, don't solve the problem.

Some open approaches should be further analysed as the BBC Open Archive. The BBC is facing a difficult challenge in negotiating rights to release content through the creative archive. There were also concerns that public institutions such as the BBC and museums may ask citizens to pay for access to digital archives, while there was feeling that tax payments should already cover this. For the BBC, payment for archiving activities was presumed to come from the reprioritization of existing funds to deliver the public value and openness the BBC is committed to in its Charter.

While the BBC through its creative archives project, and other public sector publishers are encouraging users to 'rip, mix and burn' content for transformative uses, rightsholders have similar concerns that such activity will further undermine their fight to get citizens to respect copyright and resist infringing uses. On the other side of the debate, there were concerns that with DRM and protections available for these technologies in law, the existence of the public online was severely threatened and that commercial content was 'taking over'.

There's a general consensus that both types of content should be able to exist online, however, it is recognised that there may be difficulties in teaching users to understand the difference between the two.

6.4 Outcomes

In this chapter the market of Digital Content has been analysed. In particular three aspects have been considered:

- the challenges for the Pervasive Media or Ambient Intelligent (AMI) that will embrace all aspects of ICT in the near future, at least because several entities and institution aim at this as common objectives
- the characterization of the type of content for this market with respect the one managed by DILIGENT
- the IPR protection and the possibility to have sustainability through openness (i.e. Creative Commons) or some kind of license schemas.

With respect to all these aspects DILIGENT have been compared with, we find the different results:

- DILIGENT is in line with the evolution towards pervasiveness of ICT in the environment since it aims at bringing on the eInfrastructure with all the capabilities to manage contents for any number of Virtual Communities (of any type).
- DILIGENT has no restrictions on the type of content to be managed and currently cover almost all these types with existing scenarios and partners capabilities
- DILIGENT miss to address in any form the issue of Intellectual Property Rights management. There are some security functions at level of content encryption and – of course – authentication and authorization, but no particular attention is given to supporting any license schema, payments, etc.

7 REASSESSMENT OF INTERNAL AND EXTERNAL FACTORS

In this chapter we update the critical success factors for DILIGENT and reassess the SWOT analysis which distinguishes between internal and external environments and which can potentially evolve with time. The understanding of these factors is critically relevant to ensure that the sustainability strategy of DILIGENT is chosen and adjusted properly. The consequences of this and some measure that will be taken in the next period will be discussed in the subsequent Conclusion Chapter.

7.1 Critical success factors

Based on the analysis of D4.3.1 “Market and Technology Trends Analysis” and following the observations and analysis in the previous chapters we have identified the following factors as most important in order for DILIGENT to succeed in delivering value to users:

1. **Continued funding**

The Digital Libraries market is immature as far as user requirements and product maturity is concerned. We therefore expect that achieving the vision of cost efficient and flexible Digital Libraries on Grid infrastructure will require external funding for several years to come in addition to the revenue opportunities that have been identified in D4.3.2.

2. **Delivering the right functionality at the right time**

This is a common success criterion for all product development organizations. Specifically for DILIGENT, it means the ability to continuously identify the most attractive market segments in the Digital Libraries space and focus on delivering features that tightly match user requirements.

3. **Quick wins**

Diligent must be able to demonstrate with real world examples that it is meaningful to Digital Libraries on Grid infrastructure. This factor also relates to the **recommendation 5** raised by the DILIGENT reviewers at the 1st review which highlights the importance of the sustainability in terms of impact on the communities dealing with grid and DL research and development. In Chapter 4 of this report the foreseen impact on the user communities and content providers which are represented by DILIGENT partners has been investigated in order to demonstrate the value of DILIGENT results to these target groups.

4. **Technology pragmatism.**

Due to the evolving nature of Grid technology, DILIGENT must be pragmatic in using alternative technologies for a migration path to a Grid-solution only. This factor has also been recognised by the DILIGENT reviewers’ **recommendation 3 distributed after the 1st project review** which proposes that awareness of grid initiatives other than EGEE (with alternative middleware) might mitigate technological risks. In Chapter 5 of this report other middleware have been selected and analysed in order to be aware of migration opportunities.

7.2 SWOT Analysis

A SWOT analysis has been produced in D.4.3.1 “Market and Technology Trends Analysis”. The following is an update taking into account the new findings as described in the previous chapters. In general a SWOT analysis enables the distinction between internal and external environments. Strengths and Weaknesses refer to the “Internal

environment” whereas Opportunities and Threats refer to the “External environment” and therefore involve an assessment of risks.

As the dynamics of the internal environment is quasi-static, a change of the Strengths and Weaknesses is most likely to be caused by the underlying enabling technologies. The external environment is more dynamic as it is influenced by external factors such as the market evolution. It is the scope of this Market Watch report as well as the next increment of this ongoing deliverable (D4.3.3b, M31), in the following, we update the SWOT analysis in order to facilitate a critical assessment of the upcoming business plan definition in D4.3.4.

Strengths:

- EU funding (i.e public funding)
- Cross discipline competences (i.e. Grid and Digital Libraries technologies)
- Close cooperation between technology providers and users
- Clear understanding of potential and weakness of underlying grid technology

Opportunities:

- Cost efficient enabler of Digital Libraries solutions
- Open solution to third party grid/digital library services
- Adoption of the DILIGENT developed technology in other application areas
- Ongoing digitisation trend of non-digital media may increase demand for cost-efficient Digital Libraries solutions such as provided by DILIGENT

Weaknesses:

- Research and commercial partners with conflicting objectives and expectations regarding commercialization of project results
- Grid technology limitations in enabling real-time response applications

Threats:

- Grid technology still immature for commercial exploitation, i.e. services might temporarily be unable to meet customer expectations
- Large initial investments required to establish *tota*/Grid infrastructure
- Small market for Digital Libraries outside academia

8 CONCLUSIONS

In this Market Watch deliverable findings are reported related to the monitoring of the evolution of the Digital Library (DL) and Grid markets. This report updates the DILIGENT project on current trends (technological as well as marketwise), related products, and emerging competitors. This conclusion chapter will summarise the main findings of the previous chapters and propose important next activities necessary to support the definition and refinement of the business model for DILIGENT which will be carried out in D4.3.4a "Preliminary Business Plan".

Still today, the approach of DILIGENT to develop and establish a Digital Library infrastructure which embraces the resource sharing paradigm of grid technology is **unique in the DL domain** and clearly sets DILIGENT apart within the competitor spectrum.

The comparison with existing solutions for digital libraries and scientific repositories as characterised in Chapter 3 gives clear requirements on the functional level which DILIGENT has to satisfy in order to meet and exceed user expectations. Although a technical analysis of the functional comparison is not subject of this report, there are obvious indications that DILIGENT is the only candidate to act as the digital library infrastructure which facilitates knowledge sharing and remote co-operation in e-Science which will form an integral part of the e-infrastructure for research in the future.

An initial analysis of the best practices of and **DILIGENT's impact on the user communities** which are already represented in the DILIGENT consortium has been performed (Chapter 4) in order to better understand the expectations of the user communities, identify areas of impact, and to support establishing a viable sustainability strategy. Both interview partners, ESA and SNS, as representatives of the Earth Observation Science and Humanities communities said that the ability of the DILIGENT infrastructure to allow cost and time effective on-demand creation of digital libraries and repositories will affect their usage patterns of DLs strongest.

- In addition, the Earth Observation community (represented by ESA) anticipates that DILIGENT will have enormous effect on their content management and exploitation capabilities. This will lead to technical innovation and social impact as the formerly impossible automatic creation and maintenance of environmental reports will transform these reports into powerful *living documents* to be executed on. However, this will depend on DILIGENT's technical capability to leverage the management of large and distributed VOs and the integration with existing Earth Observation services.
- For SNS (representing the Humanities community) additional value will be created because DILIGENT will facilitate the creation of large multimedia knowledge repositories which provides a platform for collaboration of scholars and students in courses and workshops.

The sustainability of impact of DILIGENT depends on the ability to demonstrate in real-world scenarios the value creation for the user communities. The importance of this fact has also been highlighted by the **recommendation 5** raised by the DILIGENT reviewers as a result of the 1st project review. The initial analysis of representatives of the two user communities has documented the areas for impact on the communities and simultaneously indicated the technical requirements to meet their expectations. In order to better understand the communities and to achieve a higher level of self-containment in the analysis the next version of this document, D4.3.3b, foreseen by M31, will be focused on the impact analysis including other members of these communities.

Even outside of the research communities which are the primary target group of DILIGENT there are **market opportunities** due to the potential of DILIGENT to help content owners in the audiovisual industry to manage challenges which arise in the ongoing paradigm shift in content distribution. As one example, RAI, as a major actor in the audiovisual sector, sees opportunities for DILIGENT to establish new and cost-effective distribution channels for educational content.

To understand the reference market for DILIGENT which goes beyond the scientific and educational domains, this report also illuminated the **Digital Content Market** at large to identify and analyse possible target areas for later DILIGENT services or products. Chapter 6 characterised the types of content in the market explicitly relating to the one managed by the project. DILIGENT is compliant with the evolution towards pervasiveness of ICT in the content environment as one of its objectives is to provide eInfrastructure with all the capabilities to manage contents for any number of Virtual Communities (of any type). Currently, there is no principal restriction on the types of content to be managed by DILIGENT – the project covers almost all referenced types with existing scenarios and partners' capabilities. Due to this, DILIGENT is in line with the evolution towards pervasiveness of ICT in the content environment.

Besides the solid technological foundation of DILIGENT, the analysis of the effects of IPR in content distribution shows that DILIGENT misses to appropriately address the issue of **Digital Rights Management** and the protection of IPR which goes beyond low-lying security functions at the level of content encryption, authentication and authorization which are supported already. If DILIGENT wants to gain the trust of content owners outside the domain of user-generated or publicly available content whose surrounding business models try to achieve sustainability through openness (i.e. Creative Commons), DILIGENT has to give particular attention to the support of arbitrary license schemas and flexible payment models for creation and distribution of digital content.

In the context of the **grid as main enabling technology** the market shows an ongoing increase of awareness which is mandatory for a later industrial take-up of DILIGENT results. The analysis demonstrates an ongoing convergence of grid-related standards towards a web-service based notion of grid services – a paradigm which is already adopted by DILIGENT since the early design phase. This overall convergence will be fostered by the recent establishment of the Open Grid Forum (OGF) which will deal with the challenges of the standardisation process. Technologywise, OGF will be one of the major communities in which DILIGENT has to discuss technical choices and submit requirements. The alliance with the grid communities is crucial to ensure impact and receive support. The update on the monitoring of existing grid middlewares shows that there is no viable commercial solution which could compete with the gLite adoption in terms of dependability and support. In addition, the start of the EGEE II project will allow for an extension of the ongoing fruitful collaboration which assures a full support of grid middleware and considerations of DILIGENT requirements.

It will be the objective of the next deliverable D4.3.4a "Preliminary Business Plan" to define a **sustainable approach to valorise DILIGENT results** in a possible commercial follow-up of the project. Although this will probably happen after a maturing of the testbed functionality of DILIGENT, it is important to understand the challenges on the path to commercialisation. As for any commercial product or service that is built on innovation it is crucial that:

- The innovation must have a visual and practical impact on the end-user experience
- The usage of the innovation must have benefits across a wide range of application scenarios

- The innovation must fundamentally change or add to what problems that can be efficiently addressed with the underlying technology
- The benefits must be validated in real world customer settings

To achieve these requirements critical success factors have been defined in Chapter 7. Most relevant in this respect is that DILIGENT is able to deliver the right functionality at the right time to the right target group. This depends on the delivery of technical functionality and a good understanding of the user communities of DILIGENT. While the solution of the technical aspects is not subject of the Exploitation workpackage – although the lack of support for DRM, license schemes, and payment models has been highlighted – it is necessary to understand the impact opportunities in the user communities. To extend the current understanding, in the next version of this document, D4.3.3b, foreseen by M31, the impact analysis will be expanded including other members of the communities included in DILIGENT.

9 APPENDICES

In this chapter information that support the reasoning made in chapter 5 on grid as enabling technology is reported. In particular:

- the EGEE project, its successor EGEE II and the foreseen setup of EGI is described (Sec. 9.1) to highlight the continuity in supporting gLite and hence DILIGENT use of this middleware;
- the other Grid middleware initiatives presented in section 5.2.1 and listed in Table 3 are further detailed (Sec 9.2). Parts of the descriptions provided have been directly taken from the projects' web pages available at the URLs listed in Table 3.
- the standard bodies (Sec. 9.3) involved at different level in the definition of standards that will have an impact on grid technology;
- the current standards (Sec. 9.4) used within most of the grid projects including DILIGENT.

9.1 Focus on EGEE

DILIGENT has chosen the EGEE project as the main middleware and infrastructure provider. DILIGENT adopts, experiments with, and extends EGEE achievements on the interest of the Digital Libraries community and users. At this purpose, a close collaboration between EGEE and DILIGENT has been established and carried on from the very beginning of the project providing requirements, resources and feedback of value to EGEE.

The success of the EGEE project strongly relies on the usage of the Grid infrastructure by various different applications. Creating a large, varied user community across Europe and beyond is therefore one of the central objectives of EGEE. By enriching the infrastructure with new requirements, services and users, DILIGENT will contribute to the uptake of the technology that will ensure an optimum long-term sustainable use of the infrastructure on a European scale.

From EGEE to EGEE-II

The Enabling Grids for E-science (EGEE) project was conceived as the first two-year phase of a four-year programme to provide a production quality Grid infrastructure across the European Research Area and beyond.

Phase two of EGEE, called the EGEE-II project, started on 1 April 2006 and builds on top of the work of the EGEE project. EGEE and EGEE-II are co-funded by the European Commission.

EGEE-II provides a 24/7 Grid Production Service for scientific research. Already serving many scientific disciplines, it aims to provide academic and industrial researchers with access to major computing resources, independent of their location. By extending EGEE, the EGEE-II project extends the boundary of Grid collaboration in Europe bringing to a larger number of partners and countries involved, an expanding infrastructure, improved middleware functionalities, a wider portfolio of application domains and a network of related projects collaborating in various areas.

A large collaboration

The EGEE-II Consortium consists of more than 90 partners from 32 countries, grouped into 13 federations and representing almost all major and national Grid efforts in Europe, as well as projects from the US and Asia. In addition, collaborations have been established with a number of related projects spun out from or affiliated with EGEE and EGEE-II. With an expanded Consortium of enthusiastic participants and a large range of

related projects, EGEE-II will be able to further develop its infrastructure into a truly pervasive global platform for e-Science.



Figure 9.1: Partner countries of the EGEE-II project (<http://www.eu-egee.org>)

An expanding Grid infrastructure

The EGEE-II project will significantly extend and consolidate the EGEE Grid infrastructure, which links national, regional and thematic Grid efforts, as well as interoperate with other Grids around the globe. The resulting high capacity, world-wide infrastructure greatly surpasses the capabilities of local clusters and individual centres, providing a unique tool for collaborative compute-intensive science ("e-Science"). Moreover, a number of related projects being submitted to FP6 calls will extend the infrastructure further, to the Mediterranean area, Baltic States, Latin America and China.



Figure 9.2: Country coverage of the EGEE middleware infrastructure including related projects (<http://www.eu-egee.org>)

The EGEE Grid consists of over 20,000 CPUs available to users 24 hours a day, 7 days a week, in addition to about 5 Petabytes (5 million Gigabytes) of storage, and maintains 30,000 concurrent jobs on average. This is the largest Grid infrastructure in the world. Having such resources available changes the way scientific research takes place.

Improved Grid middleware functionality

Originally, EGEE used middleware based on work from its predecessor, the European DataGrid (EDG) project, later developed into the LCG middleware stack, which was used on the EGEE infrastructure early in the project. In parallel, EGEE has developed and re-engineered most of this middleware stack into a new middleware solution, gLite, now being deployed on the preproduction service. The gLite stack combines low level core middleware with a range of higher level services. Distributed under a business friendly open source license, gLite integrates components from the best of current middleware projects, such as Condor and the Globus Toolkit, as well as components developed for the LCG project. The product is a best-of-breed, low level middleware solution, compatible with schedulers such as PBS, Condor and LSF, built with interoperability in mind and providing foundation services that facilitate the building of Grid applications from all fields. gLite Grid middleware, described in detail in section 4.3, adds new features in all areas of the software stack. In particular it features better security, better interfaces for data management and job submission, a re-factored information system, and many other improvements that make gLite easy to use as well as effective. gLite improves the infrastructure and forms an even more dependable and scalable resource, able to meet the needs of a large, diverse e-Science user community.

A wide portfolio of application domains

Expanding from originally two scientific fields, high energy physics and life sciences, EGEE now integrates applications from many other scientific fields. Researchers in academia and industry already benefit from the EGEE e-Infrastructure, which simultaneously supports many applications from diverse scientific areas, providing a common pool of resources, independent of geographic location, with round-the-clock access to major storage, compute and networking facilities. Generally, the EGEE Grid infrastructure is ideal for any scientific research especially where the time and resources needed for running the applications are considered impractical when using traditional IT infrastructures. Today more than 60 VOs and many more applications run on the EGEE infrastructure. EGEE-II will expand the portfolio of supported applications, including already 9 scientific domains: High Energy Physics, Biomedicine, Earth Science, Astrophysics, Computational Chemistry, Fusion, Geophysics, Finance, Multimedia.

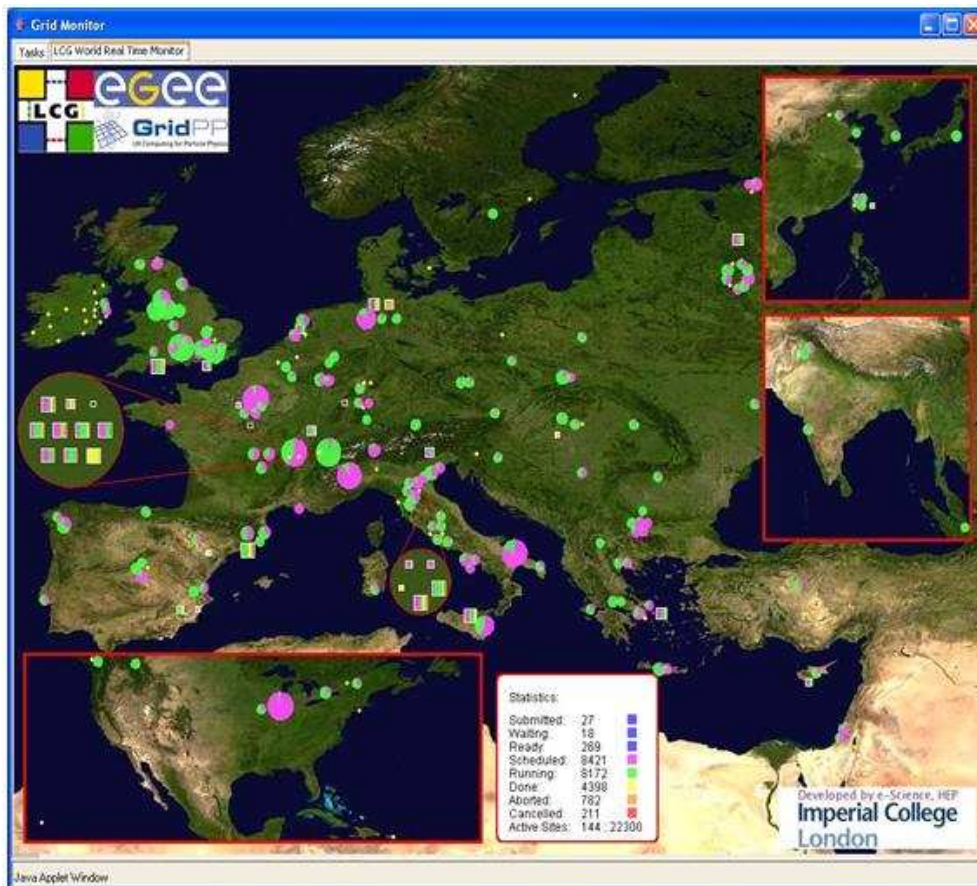


Figure 9.3: EGEE infrastructure monitoring (<http://www.eu-egee.org>)

A network of related projects

EGEE also works alongside several other major European projects. These include infrastructure projects expanding the geographical reach of the European Grid, applications projects and support projects.

Infrastructure projects have been extending the reach of the EGEE infrastructure both within Europe and outside. So far, EGEE has been involved in projects whose scope extends to South-Eastern Europe, the Baltic region, Latin America, the Mediterranean region, China and India.

Application projects build on EGEE's middleware, gLite, to provide the specific functionality required for their particular research domain. Their feedback also helps EGEE to refine its operations be more responsive to users' needs.

Finally, there are a number of support projects which enable EGEE to further improve the EGEE's infrastructure and gLite middleware. EGEE is working with projects from many areas, including computer security, e-Infrastructure policy, education, dissemination and software quality assurance.

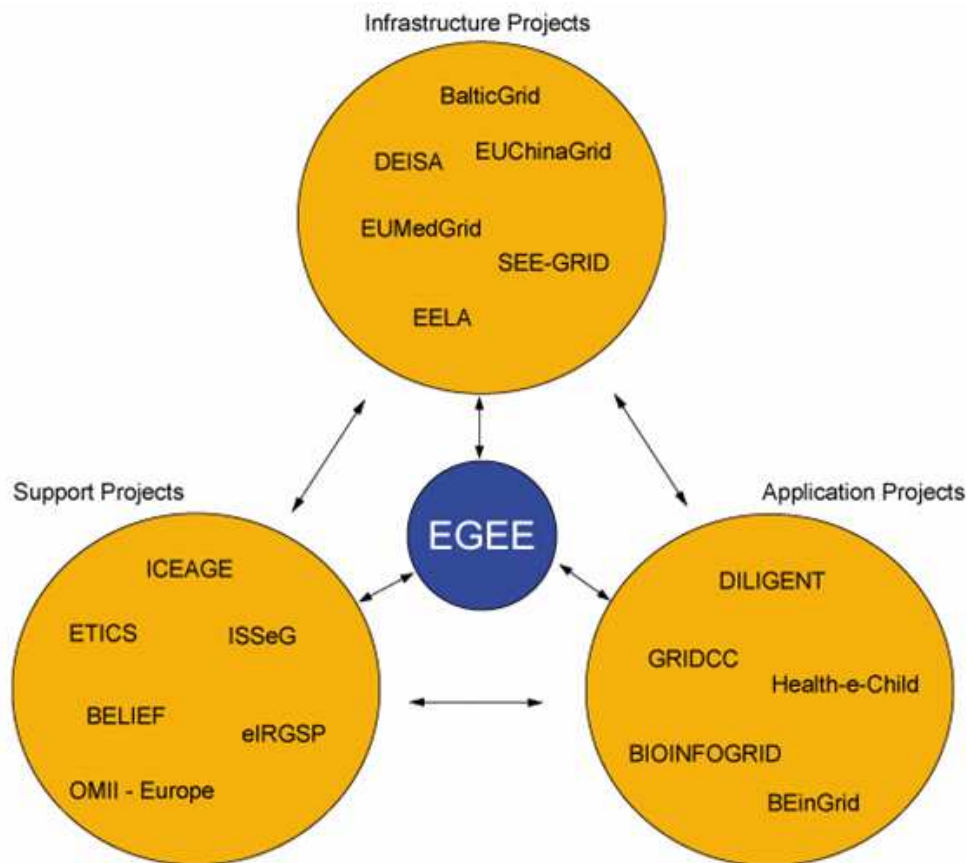


Figure 9.4: EGEE collaboration with Related Projects (<http://www.eu-egee.org>)

Collaboration with the industry

EGEE has established links with many companies as partners, users, resource providers, affiliates. All industrial partners have actively contributed to the promotion of EGEE in business sectors. The relative maturity of EGEE now makes it possible to further pursue commercial exploitation of its products and services. EGEE-II marks a significant increase in the priority and resources invested by the project in industrial exploitation.

The Industry Forum aims to raise awareness of EGEE within industry and encourage businesses to participate in the project and to use its results. This is achieved by making direct contact with industry, liaising between the project partners and industry in order to ensure businesses get what they need in terms of standards and functionality and ensuring that the project benefits from the practical experience of businesses. The members of the EGEE Industry Forum are companies of all sizes who have their core business or a part of their business within Europe. They are manufacturers, software vendors and service providers as well as end users.

Nevertheless, it has to be taken in consideration that there are limits on the commercial exploitation of the EGEE infrastructure:

1. GÉANT/NREN dispose restrictions on the network usage; more prominently the infrastructure cannot offer a commercial Grid service to companies
2. Many resource providers are not in a position to charge for the usage of the resources they share

3. The short-term funding cycle of EGEE determines uncertainty of long-term sustainability of the infrastructure itself

Hence EGEE focuses on *users*, at a pre-competitive stage, *technology providers* (links with IT vendors, service providers and system integrators) and *transfer of technology* to create commercial products and services. gLite middleware is distributed under a business-friendly open source license and Grid operations procedures, practices & tools are implemented as best-of-breed examples.

Business partners involved in EGEE are DATAMAT (Italy), NICE (Italy) and CS-SI (France). Industry applications are implemented by EGEODE and the OpenPlast project. Additionally, EGEE has regular exchange with the Business Experiments in Grid project (BEinGRID, <http://www.beingrid.com>), coordinated by ATOS Origin (Spain). BEinGRID put in place a private Grid infrastructure. EGEE partners support the work of BEinGRID which aims at evaluating gLite usage in commercial contexts. At present there are 3 experiments based on gLite middleware: financial portfolio management (Financial sector), insurance risk management (Insurance sector) and seismic imaging & reservoir simulation (Industry sector Oil & Gas) plus others under consideration.

Finally, EGEE accepts Business Associates, which are offered the opportunity to sponsor work on joint-interest subjects. The benefits for them are: transfer of technical know-how, early access to documentation and key project deliverables, sponsorship and visibility at EGEE events, contribution to the programme of industrial events as members of the Industrial Forum Steering Committee, participation early-on in the process to define a sustainable infrastructure after EGEE-II.

A sustainability strategy for the EGEE infrastructure

The potential of Grid computing is to be fully realised in the years to come. To exploit fully the benefits that Grid computing can offer to European science and industry, it is necessary to ensure the continued and sustainable long-term availability of such computing facilities.

Efforts such as EGEE which are funded on two-year funding cycles are unable to provide sufficient guarantees for the long term availability of production-quality services and risk dispersing the skill base that the FP5 and FP6 Grid projects have built up.

Comment [fm1]: new after TCom meeting in Rome

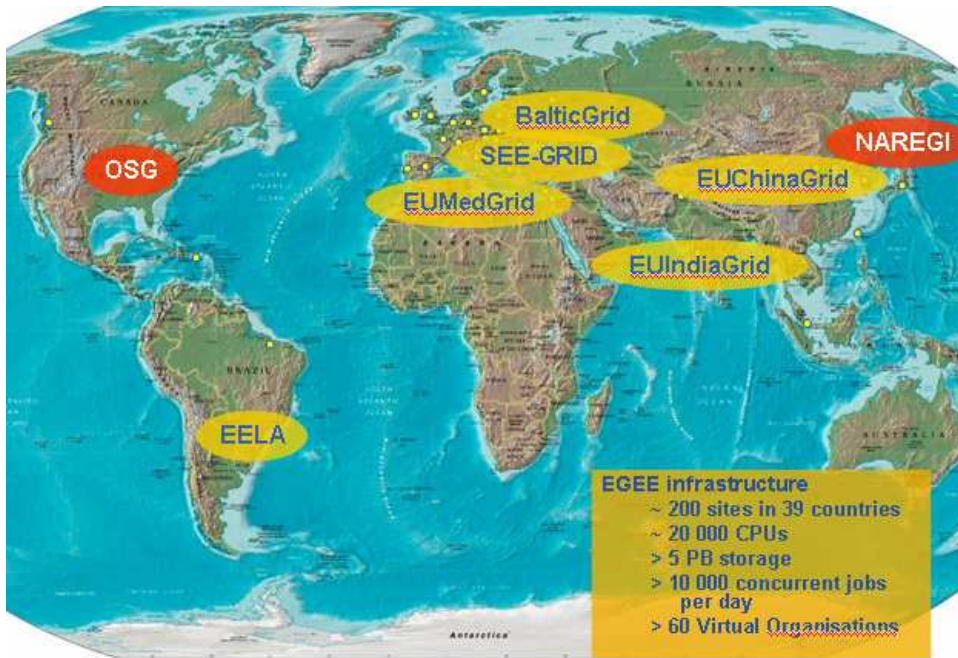


Figure 4.1: The EGEE and related-project e-infrastructure (<http://www.eu-egee.org>)

Many countries now have National Grid Initiatives (NGIs) which mobilise national funding and resources, help to define international standards and policies, and operate mainly application-independent national e-infrastructure.

In considering new models for the future of the European e-Science infrastructure, the requirements of a large number of stakeholders must be taken into consideration. These include: national/EU/funding agencies, National Grid Initiatives, industry, resource providers, existing Grid projects, user communities, policy bodies and the public.

The need for a sustainable e-infrastructure has already been recognised¹² and a model of an e-Infrastructure for Europe is currently being defined and proposed under the name of European Grid Infrastructure (EGI).

The mission of EGI includes co-ordination of a production Grid infrastructure for European Research Area, interoperation with e-Infrastructure projects around the globe and contribution to Grid standardisation and policy efforts. This service is intended to support applications from diverse communities (Astrophysics, Computational Chemistry, Earth Sciences, Finance, Fusion, Geophysics, High Energy Physics, Life Sciences, Material Sciences, Multimedia, and more). In addition, EGI will forge links with the full spectrum of interested business partners and disseminate knowledge about the Grid through training.

¹² Recommendations of the Global Science Forum of the Organisation for Economic Co-operation and Development in the report on Grids and Basic research Programmes, Sydney September 25-27 2005, <http://www.oecd.org/dataoecd/30/36/36213997.pdf>
Recommendation of the e-Infrastructures Reflection Group (www.e-irg.org) in the White Paper: <http://www.e-irg.org/publ/2005-Luxembourg-eIRG-whitepaper.pdf>
Establishing an European Grid Organisation (EGO) paper: <http://www.e-irg.org/meetings/2005-UK/050617-EGO-position-paper.pdf>

The proposed structure is based on a federated model bringing together National Grid Initiatives (NGIs) to build a European organisation. Current EGEE-II federations are expected to evolve into NGIs. Each NGI is a national body recognised at the national level which mobilises national funding and resources, contributes and adheres to international standards and policies, operates the national e-Infrastructure, is application independent, open to new user communities and resource providers.



Figure 4.2: The EGEE federations. (<http://www.eu-egee.org>)

Many European National Grid Projects are already active or planned to start soon¹³.

Key services such as coordination of infrastructure operations, middleware testing and certification, application support, dissemination and outreach, training could be part of a central organisation coordinated with the NGIs.

EGI will cooperate with super-computer centres to the user communities who want to see a unified e-Infrastructure providing access to both capability (super-computer centres) and capacity (PC cluster-based centres). Several communities (Fusion, Life sciences etc.) have a need for both types of facilities and want to be able to move applications and data between them. But EGI will not be responsible for the operations, funding or management of super-computer centres.

In this plan a business model for how industry can commercially exploit the research infrastructure managed by EGI/NGIs is still unclear. It should be taken into consideration that public funding should not be used to compete with commercial service providers; therefore the infrastructure can be used by companies to do research at a pre-competitive stage and encourage transfer of technology. This transfer of technology from research to industry would happen by adoption/internalisation of EGI/NGI backed products and services (e.g. middleware, operations procedures/techniques). Current e-Science applications could lead to e-Business applications; experiences in running a distributed infrastructure would lead to multi-site corporate usage. Several examples of technology transfer and cooperation with Industry already exist in EGEE. EGI/NGIs could subcontract infrastructure support to industry and make use of commercial software as standards evolve.

¹³ (*Italics indicate in planning stage*): Austria – AustrianGrid, Belgium – BEGrid, Bulgaria – BgGrid, Croatia – CRO-GRID, Cyprus – CyGrid, Czech Republic- METACentre, Denmark, Estonia – EstoniaGrid, Finland, France – ICAR, Germany – D-GRID, Greece – HellasGrid, Hungary, Ireland - Grid-Ireland, Israel - Israel Academic Grid, *Italy – planne*, *Latvia – Latvian Gri*, Lithuania – LitGrid, Netherlands – DutchGrid, Norway – NorGrid, *Poland – Pioneer*, *Portugal – launched Apr'06*, Romania – RoGrid, Serbia – AEGIS, *Slovakia*, Slovenia - SiGNET, *Spain – IBERgrid*, Sweden – SweGrid, *Switzerland – SwissGrid*, Turkey – TR-Grid, *Ukraine – UGrid*, United Kingdom - eScience

Given the existence of such an e-Infrastructure, resources centres, managed by the NGIs and coordinated by EGI, can be established to create shared pool of resources (CPU, disk and data curation) independent of funding for specific user communities. A business model with pay-per-usage could be adopted to cover operational and depreciation costs and would create/test example business models for potential future commercial supply and/or exploitation.

EGI is expected to be an organisation with its own legal identity where NGIs are the stakeholders and form the governing council. EGI would be subject to annual reviews by independent experts nominated by the EU.

The EGI model proposed therefore commits the National Grid Initiatives in order to ensure a sustainable e-Infrastructure for research and help maintain Europe's leading position.

EGEE-II in numbers

the infrastructure	
~200	sites running EGEE Grid middleware
20.000+	CPUs
32	countries involved and more through EGEE related projects
the collaboration	
91	partners
48	non contracting partners
139	Participants including US and Asia
32	countries involved
13	federations
the user community and usage	
9	scientific domains
60+	VOs using the EGEE production service
~30.000	jobs per day

Table 6: EGEE II in numbers

Websites

EGEE homepage <http://www.eu-egee.org>

gLite website <http://www.glite.org>

Try the Grid <http://public.eu-egee.org/test/>

gLite license <http://public.eu-egee.org/license/license2.html>

9.2 Other Grid middleware initiatives

The Grid middleware initiatives presented in Table 3 are further detailed in this section: each project is described by a text part and by the information on its license type.

Parts of the descriptions provided below have been directly taken from the projects' web pages available at the URLs listed in Table 3.

Deisa - Distributed European Infrastructure for Supercomputing Applications

Commercial: No

The main objective of the DEISA Research Infrastructure is to advance computational sciences in Europe. The DEISA Extreme Computing Initiative (DECI) has been launched in May 2005 by the DEISA Consortium to enhance its impact on science and technology: its main purpose is to enable a number of challenging applications in all areas of science and technology. Basic requirement is that these applications must deal with complex, demanding and innovative simulations that would not be possible without the DEISA infrastructure.

The DEISA applications are expected to have requirements that cannot be fulfilled by the national services alone. A few examples of application profiles and use cases that are well adapted to the present status of the DEISA supercomputing Grid are the following:

- International collaborations involving scientific teams that access the nodes of the AIX super-cluster in different countries
- Extreme computing demands for challenging projects requiring a dominant fraction of a single supercomputer
- Workflow applications involving at least two platforms
- Coupled applications involving more than one platform

The DEISA Consortium provides support to the users: the Applications Task Force (ATASKF) is a team of leading experts in high performance and Grid computing whose major objective is to provide the consultancy needed to enable the user's adoption of the DEISA research infrastructure. The ATASKF helps users to design and adapt applications to the DEISA infrastructure: currently its activities can be grouped in Hyper-scaling of parallel applications, Design of workflow applications, Design of coupled and grid applications, Data intensive applications.

EnFuzion

Commercial: Yes

EnFuzion offers application developers a simple and powerful scripting interface to grid-enable, parameterize and distribute their applications to run on clusters and grids. A standard HTTP or a network-based API can be used to control EnFuzion from an application to gain additional flexibility and improve ease of use. EnFuzion can complement, co-exist and co-operate with application-specific job schedulers in heterogeneous environments. Axceleon works closely with leading application vendors to offer an ever growing list of industry-specific solutions.

Main EnFuzion features are:

- Virtual supercomputer based on all the computing resources of a corporate network
- Multi-user priorities to process applications concurrently
- Resource management, only resources satisfying job requirements are being used to execute the job

GFarm

Commercial: No

Gfarm file system is a next-generation network shared file system, which will be an alternative solution of NFS, and will meet a demand for much larger, much reliable, and much faster file system.

Gfarm is a reference implementation of the Grid Datafarm architecture designed for global petascale data-intensive computing. It provides Gfarm Grid file system that is a shared file

system in cluster or Grid that can scale up to petascale storage, and realize scalable I/O bandwidth and scalable parallel processing.

Gfarm Grid file system is a virtual file system that integrates local disks of compute/filesystem nodes. It consists of:

- many compute/filesystem nodes, and
- Gfarm metadata server node

On each compute/filesystem node, the Gfarm file system daemon (gfsd) is running to facilitate remote file operations with access control in the Gfarm filesystem as well as file replication, fast invocation, and node resource status monitoring.

Gfarm metadata server node manages Gfarm filesystem metadata and parallel process information, on which the Gfarm job manager (gfmd), and filesystem metadata server (slapd) are running.

GPT

Commercial: No (open source)

GPT is a portable software packaging system that keeps track of the software installed on a system and manages dependencies between packages. GPT adds metadata to standard tar.gz files, putting more "intelligence" into the build/install/config/uninstall process. GPT uses the Perl language.

GridAssist

Commercial: No (open source)

GridAssist addresses the issue of executing workflows across Virtual Organizations. GridAssist is a workflow management tool for computational Grids: GridAssist is a software framework that provides benefits of workflow computing in a Computational Grid environment to applications that are not intrinsically Grid-aware, for users that do not have to be experts on Grid technology.

GridAssist is designed for creating and running scientific data processing jobs efficiently on resources provided by computational Grids. The tool hides the technical details where and how to run the applications in the workflow and takes care of the transfer of the data between the resources. The tool provides access to Grids directly from the users desktop.

Gridblocks

Commercial: No (OSI compliant BSD-license)

The Gridblocks mission is to build a grid application framework with open, standardized interfaces facilitating the creation of end user services via easy-to-use building blocks in distributed environment.

Gridblocks in details:

The security services are the backbone that are found from each Grid block. GB Core is based on GSI & EDG Security work. It will be extended with open standards of Liberty Alliance to enable easy path to creation of virtual organizations.

Apart from interfacing solely on Globus Toolkit GridBlocks provides a alternative Grid environment, that does not require much to get started. GB Agent is an all Java approach to job submission, data retrieval and results analysis presentation with strong Agent paradigma.

Portal is the user-friendly interface to Grid applications. GB Portal builds on common XML presentation layer that provides a set of value added services and easy connectivity.

Portal aims to combine personalizable dynamic views to command Globus and GridBlocks services in unified look and feel in variety of content forms: XHTML, WML and PDF for example.

P2P is promising technology to connect limited mobile devices to Grid. Continuous Net connection and the absence of firewalls make P2P an ideal fit for future Mobile devices. GB Peer is another alternative access point to GridBlocks services.

Server components deal with providing high availability (HA) services that can optionally be plugged in with other Grid blocks. J2EE application server clusters are de-facto Java platforms for hosting Web Services and future Grid Services (Globus 3).

GridFTP

Commercial: No (open source)

GridFTP (current version GT4.0/NMI-R7) is a high-performance, secure, reliable data transfer protocol optimized for high-bandwidth wide-area networks. It is based upon the Internet FTP protocol, and it implements extensions for high-performance operation that were either already specified in the FTP specification but not commonly implemented or that were proposed as extensions by our team. The current GridFTP protocol specification is now a "proposed recommendation" document in the Global Grid Forum (GFD-R-P.020).

GridFTP uses basic Grid security on both control (command) and data channels. Other features include multiple data channels for parallel transfers, partial file transfers, third-party (direct server-to-server) transfers, reusable data channels, and command pipelining.

Associated standards for GridFTP:

- RFC 959 Base FTP protocol
- RFC 2228 gssapi security extensions for FTP RFC 2389 FEAT, OPTS, etc.
- extensions to FTP (IETF FTP Working group draft) for structured directory listings, SIZE, MDTM commands.
- GFD.020 GridFTP extensions

GridICE

Commercial: No (open source)

GridICE is a distributed monitoring tool designed for Grid systems. It promotes the adoption of de-facto standard Grid Information Service interfaces, protocols and data models. Further, different aggregations and partitions of monitoring data are provided based on the specific needs of different user categories (VO, GOC). It is possible to verify the composition of virtual pools or to sketch the sources of problems.

GridICE integrates with local monitoring systems and offers a standard interface for publishing monitoring data at the Grid level (the default fabric monitoring tool is Lemon from CERN). The set of attributes that are measured is an extension of the GLUE Schema v.1.1. The distribution of monitoring data follows a two-level hierarchy (local site collection, grid-wide collection). The global monitoring information can be accessed in different ways: web-based interface offering both textual a graphical representation, XML representation over HTTP for application consumption and publish/subscribe for the notification of events of interest.

GridICE was started in 2003 during the European DataTAG project and is being evolved in the framework of the EGEE project. Among the several future activities, it considers the integration with the metering service of the DGAS accounting system, fine-grain access to monitoring data using Grid-based credentials and the extension of sensors in order to

offer an interface for the push of monitoring data by the local site dealing with privacy concerns.

Main GridICE's features:

- automatic discovery of new resources to be monitored through the Grid Information Service
- notification service
- complete set of monitored metrics, from host-related to Grid service related characteristics
- supports and extends the GLUE Schema/support for the following batch systems: OpenPBS, Torque, LSF
- each view of the web-based interface offers the same data in XML format
- integrated with network-related infrastructure for monitoring the connectivity of a Grid
- maintain history of metrics and support a per-attribute threshold-based mechanism for data reduction

Grid MP

Commercial: Yes

Grid MP is a commercial distributed computing software package developed and sold by United Devices. It was formerly known as the MetaProcessor prior to the release of version 4.0. Main Grid MP's features are:

- job scheduling with prioritization
- user security restrictions
- selective application exclusion
- user-activity detection
- time-of-day execution controls

Grid MP can be used to manage computational devices consisting of corporate desktop PCs, departmental servers, or dedicated cluster nodes.

Additional components are:

MGSI: MAPI Grid Services Interface, or simply MGSI, offers a programmatic web service API (via SOAP and XML-RPC protocols over HTTP). It enables developers of back-end application services to access and manipulate objects within the system. Access to the API and all objects is access controlled and security restricted on a per-object basis. Since MGSI is a web service protocol, any programming language that has a SOAP or XML-RPC library available can be used to interface with it, although commonly C++, Java, Perl, and PHP are used.

Management Console: a web-based MP Management Console, or simply MPMC, provides administrators with a simplified and easy-to-use interface to monitor system activity, control security settings, and manage system objects. Of technical interest is that the MPMC is written in the PHP programming language, and uses the MGSI web service for all of its interactions with the system.

MP Agent: the MP Agent is the software agent that must be installed on each computer that will participate in a Grid MP installation by running jobs. Once the MP Agent is installed on a computer, it is officially recognized by the Grid MP as a Device.

GSI Plugin for gSOAP

Commercial: No

The Grid Security Infrastructure Plugin for gSOAP is an open source solution to the problem of securing web services in a grid environment. It provides a gSOAP plugin for building secure web services using the Globus Toolkit GSI infrastructure: with this package the HTTPG (HTTP over GSI) protocol to develop GSI enabled web services and clients can be used. The main features of this plugin can be summarized as follows:

- Based on the GSS API for improved performances
- Extensive error reporting related to GSS functions used
- Debugging framework
- Support for both IPv4 and IPv6
- Support for development of both web services and clients
- Support for mutual authentication
- Support for authorization
- Support for delegation of credentials
- Support for connection caching
- Support for timeout management
- Support for automatic client-side credential renewal.

MS .NET Grid

Commercial: Yes

This collaboration between Microsoft and NeSC (represented in this project by EPCC) has the goal of providing a practical demonstration to the UK e-Science community of the applicability of Microsoft .NET technologies to the hosting, development and deployment of Grid Services. A complementary goal is facilitating understanding of the Grid and e-Science within Microsoft.

These goals have been realised by achieving the following objectives:

- Developing an implementation of the Open Grid Services Infrastructure (OGSI) using .NET technologies
- Developing a suite of Grid Service demonstrators that can be deployed under this .NET OGSI implementation
- Developing training courses and materials to educate and inform the UK e-Science community about .NET and its applicability to Grid applications

NMI

Commercial: No

The National Science Foundation Middleware Initiative (NMI) addresses a critical need for software infrastructure to support scientific and engineering research. Begun in late 2001, NMI funds the design, development, testing, and deployment of middleware, a key enabling technology upon which customized applications are built. Specialized NMI teams are defining open-source, open-architecture standards that are creating important new avenues of on-line collaboration and resource sharing. In addition to the production-quality software and implementation standards created by those large systems-integration teams, NMI funds smaller projects that focus on experimental middleware applications.

In fall 2003, NMI extended its two original systems-integration projects, the GRIDS (Grid Research Integration Deployment and Support) Center and the Enterprise and Desktop Integration Technologies (EDIT) consortium. At the same time, a pair of additional teams

joined NMI: the Open Grid Computing Environment (OGCE) team, which develops portals for ubiquitous, browser-based access to Grid resources, and the Common Instrument Middleware Architecture (CIMA) team, which develops tools that ease the Grid-enablement of scientific instrumentation.

OMII – Open Middleware Infrastructure Institute

Commercial: No

OMII (current version 2.3.3) provides a web service infrastructure for building grid applications. The OMII platform focuses on the needs of distinct yet important stakeholders within Grid computing: the Service Provider and the Client.

OMII-UK is funded by EPSRC through the UK e-Science Core programme. It is a collaboration between the School of Electronics and Computer Science at the University of Southampton, the OGSA-DAI project at the National e-Science Centre and EPCC, and the *myGrid* project at the School of Computer Science at the University of Manchester.

OSG – Open Science Grid

Commercial: No

The Open Science Grid is a US grid computing infrastructure that supports scientific computing via an open collaboration of science researchers, software developers and computing, storage and network providers. This initiative is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science.

The OSG Consortium builds and operates the OSG to give scientists from many fields of application access to shared resources worldwide: in this way OSG brings resources and researchers from universities and national laboratories together and cooperating with other national and international infrastructures. The OSG infrastructure can accommodate the widest practical range of users of current Grid technologies, in a context which maximizes the future convergence of those users to greater commonality in technology choices.

The infrastructure spans multiple Grids and its production quality, scale and internal consistency, its broad scope and the diversity of its client communities lead to additional requirements in support of sustained and robust operations.

The OSG Twiki available at <http://osg.ivdgl.org/twiki/bin/view> represents a valid help to users who want to get started with this infrastructure.

pyGMA

Commercial: No

The pyGMA is an implementation of the Grid Monitoring Architecture (GMA) Producer, Consumer, and Directory Service (or Registry) Web-Services SOAP interfaces in Python. It uses the ZSI SOAP library to aid with serialization and deserialization of messages. It is not a monitoring system but a framework that handles the SOAP communications between the monitoring components defined by the Global Grid Forum. It is up to the user of pyGMA to "fill in" the useful work that will be done in response to remote queries, subscriptions, and notifications. Therefore, the target "user" of pyGMA is really a developer that wants to connect existing or newly created monitoring components into a GMA-compatible framework. Although this component may be useful for testing or small-scale deployment, it is not intended to replace a distributed directory service.

TUPELO

Commercial: No

The Open Grid Computing Environments collaboration develops and integrates Grid portal tools and services to support scientific gateways. The OGCE team spans several different core projects including Tupelo (<http://www.nsf-middleware.org/Lists/NMIR9/ogce.aspx>).

Tupelo is a Semantic Grid-based data and metadata system. Tupelo is a data and metadata archiving system that supports object-oriented metadata schemas based on RDF-OWL (<http://www.w3.org/TR/owl-features/>), logical file naming, version and access control, and other Grid services (<http://www.globus.org/ogsa/>). Stable releases of Tupelo services, client APIs, and documentation are available from <http://dit.ncsa.uiuc.edu/wiki/index.php>. Tupelo portal clients are currently under redevelopment for closer integration with OGCE tools.

The main features of TUPELO are listed below:

- Grid Service interfaces to metadata and data services
- First-class metadata schemas for "self-describing" metadata (a.k.a. ontologies)
- Support for RDF-OWL schemas including classes, typed slots, and multiple inheritance
- Version control and transactional atomicity on all operations
- GSI authentication (<http://www-unix.globus.org/toolkit/docs/3.2/gsi/>) and object-level access control - Plug-in API's for adding new storage resource types, data transport protocols, RPC frameworks, and metadata serialization methods.

Tupelo is designed for archiving large-scale, complex scientific data and metadata collections. It is also suitable for more conventional digital libraries containing Dublin Core (<http://dublincore.org/>) or other standard digital library metadata schemas. Its OWL-based metadata framework can support a wide variety of schemas.

VDT – Virtual Data Toolkit

Commercial: No

VDT's main goal is to make it as easy as possible for users to deploy, maintain and use grid middleware. VDT is an ensemble of grid middleware that after a proper configuration makes the access to grid resources and the ability to provide others with those resources possible. In particular, VDT specifically contains three kinds of middleware:

- Basic Grid services: [Condor-G](http://www.cs.wisc.edu/condor/) (<http://www.cs.wisc.edu/condor/>) and [Globus](http://www.globus.org/) (<http://www.globus.org/>).
- Virtual data tools: This includes software tools developed by the [GriPhyN project](http://www.griphyn.org/) (<http://www.griphyn.org/>) to work with virtual data, specifically the [Virtual Data System](http://www.griphyn.org/chimera/) (<http://www.griphyn.org/chimera/>).
- Utilities: This includes a wide variety of tools, such as [GSI-Enabled OpenSSH](#), software to update GSI certificate revocation lists, and monitoring software like [MonALISA](#).

VDT supports among the others the [LHC Computing Grid Project](http://lcg.web.cern.ch/LCG/) (<http://lcg.web.cern.ch/LCG/>) and the [Particle Physics Data Grid](http://www.ppdg.net/) (<http://www.ppdg.net/>).

Xgrid

Commercial: Yes

The Apple's Xgrid framework for grid introduces a new technology solution for loosely coupled, distributed computation. Xgrid has been widely promoted as an extremely usable

solution for less technical user communities and challenges the systems management paradigm incumbent in many computational grid solutions currently deployed. A particular point to note is that the simple Xgrid framework has the potential to fundamentally change the delineation between grid users and grid maintainers, and as such to promote new types of research enabled by a self-sustaining model for managing computational grid infrastructure.

A range of third party components which can be used to extend the Xgrid framework in directions more amenable to the types of production environments currently occupied by solutions such as the widely used Globus toolkit.

Xgrid was first introduced by Apple in January 2004 as a Technology Preview (TP1). Xgrid TP1 was considered as a proof of concept, and was not designed for production applications owing to reliability, security and scalability issues. It was designed to draw feedback from early adopters as to the viability of an Apple grid computing product. Xgrid Technology preview (TP2) was released in November 2004, retaining most of the functionality of TP1, but with some underlying CLI and data format changes. This was a very widely adopted release, Xgrid based computational environments were deployed for production use, and third party components began to emerge. Xgrid 1.0 was released with Mac OS X 10.4 'Tiger' in April 2005. 1.0 introduced a significant number of changes. Xgrid 1.0 included a dependency on Mac OS X Server (TP1 and TP2 did not require a server grade operating system).

The Apple Xgrid architecture is a standard three tier architecture consisting of a Client, Controller and Agent. We will review each of these tiers in turn. The Controller, Agent and Client can all exist on a single machine, although in practice, these are more typically distributed.

9.3 Grid standardisation bodies

The standardization effort that is leading the future of Grid computing is being carried out by several groups, consortiums, communities and forums. Of all consortiums currently working in Grid standards, some are already established groups now adapting their efforts in this new direction, while others are recently formed groups, fully Grid oriented. The most relevant and well known are :

9.3.1 Distributed Management Task Force (DMTF)

The DMTF is an industry-based organization founded in 1992 to develop management standards and integration technologies for enterprise and Internet environments. DMTF technologies include the Common Information Model and Web-Based Enterprise Management. The DMTF formed an alliance with the GGF in 2003 for the purpose of building a unified approach to the provisioning, sharing, and management of Grid resources and technologies.

9.3.2 Internet Engineering Task Force (IETF)

The IETF is a large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. It is open to any interested individual. The actual technical work of the IETF is done in its working groups, which are organized by topic into several areas:

- Applications

- General
- Internet
- Operations and Management
- Real-time Applications and Infrastructure
- Routing
- Security
- Transport

IETF was founded in 1986 and has started working in the Grid domain more recently, mainly involved in the definition of security certificates.

9.3.3 Global Grid Forum (GGF)

The Global Grid Forum is the primary standards-setting body for the Grid. GGF is a community-initiated forum of thousands of individuals from industry and research leading the global standardization effort for Grid computing. It was formed in 1998 from the merger of the Grid Forum in North America, the Asia-Pacific Grid community, and the European Grid Forum (eGrid).

GGF's primary objectives are to promote and support the development, deployment, and implementation of Grid technologies and applications via the creation and documentation of "best practices", technical specifications, user experiences, and implementation guidelines. In a process similar to that used for Internet standards, the GGF creates four types of documents that provide information to the Grid community:

- Informational – a useful idea or set of ideas
- Experimental – useful experiments
- Community practice – common practices or processes that influence the community
- Recommendations – specifications

GGF efforts are also aimed at the development of a broadly based integrated Grid Architecture that can serve to guide the research, development, and deployment activities of the emerging Grid communities. Defining such architecture will advance the Grid agenda through the broad deployment and adoption of fundamental basic services and by sharing code among different applications with common requirements.

The GGF currently divides its efforts among eight standards function groups:

- Infrastructure
- Data
- Compute
- Architecture
- Applications
- Management
- Security
- Liaison

Joining a GGF working group involves simply subscribing to its e-mail list. The project members, meeting agendas, and work progress are all posted online.

9.3.4 Organization for the Advancement of Structured Information Standards (OASIS)

OASIS is a non profit international consortium that drives the development, convergence, and adoption of industry standards for e-business. OASIS was founded in 1993 as SGML Open and changed its name in 1998 to reflect its expanded technical scope such as the work in Grid computing for the definition of components as part of the WSRF framework.

OASIS produces Web services standards that focus primarily on higher-level functionality such as security, authentication, registries, business process execution, and reliable messaging. This includes developing standards such as those related to the Extensible Markup Language (XML) and the Universal Description, Discovery and Integration (UDDI) service.

Participants in OASIS can be either unaffiliated individuals or member-company employees. At least three organizations must implement a standard before OASIS will approve it.

9.3.5 Web Services Interoperability (WS-I)

WS-I is an open industry organization chartered to promote Web services interoperability across platforms, operating systems and programming languages. It was formed in 2002.

The organization's diverse community of Web services leaders helps customers to develop interoperable Web services by providing guidance, recommended practices and supporting resources. All companies interested in promoting Web services interoperability are encouraged to join the effort. Specifically, WS-I creates, promotes and supports generic protocols for the interoperable exchange of messages between Web services. In this context, "generic protocols" are protocols that are independent of any action indicated by a message, other than those actions necessary for its secure, reliable and efficient delivery, and "interoperable" means suitable for multiple operating systems and multiple programming languages.

Its role is to integrate existing standards rather than create new specifications, providing a set of rules for integrating different service implementations with a minimum number of features that impede compatibility.

9.3.6 World Wide Web Consortium (W3C)

The World Wide Web Consortium is an international consortium that works for the development of Web standards. W3C has created in 1994 by Tim Berners-Lee to promote common and interoperable protocols.

W3C's mission is to lead the World Wide Web to its full potential by developing protocols and guidelines that ensure long-term growth for the Web. The World Wide Web Consortium (W3C) has been highly successful at promoting a number of proposed standards in various stages of development that are commonly referred to as the "Web services" specifications. The bases for these standards are the HTTP protocol, the XML encoding format, the SOAP remote procedure call mechanism, and the WSDL language.

W3C members are organizations that typically invest significant resources in Web technologies. OASIS is a member, and the W3C has partnered with the GGF in the Web services standards area.

9.3.7 Others

There are also other less important groups going an important work for the future of Grid computing. Even if not involved in the definition of standards these groups help in the definition of grid solutions. Others work in more specific domains. An example is the

Liberty Alliance, an international alliance of companies, non-profit groups, and government organizations that intends to develop an open standard for federated identity management to address technical, business, and policy challenges surrounding identity. Finally, and in a different perspective, groups like the **e-Infrastructure Reflection Group (eIRG)** are working in the establishment of high level guidelines for Grid computing. The main objective of the eIRG is to support at the political, advisory and monitoring level, the creation of a policy and administrative framework for the easy and cost-effective shared use of electronic resources in Europe (focusing on Grid-computing, data storage, and networking resources) across technological, administrative and national domains. The eIRG members are official delegates from the ministries of Education of the various European countries plus some EC officials. Every six months they produce a document with their recommendation in several areas but always with the main goal of building a healthy European eInfrastructure.

The main objective of most of these standardization groups is to produce grid standards and promote its acceptance by the several interested communities. However, these groups have different backgrounds and visions. This heterogeneity leads to the co-existence of many standards that answer the requirements of many different grid applications developers. Some standards can be considered as more mature and accepted while others are still emerging. The first ones are normally older standards, heavily tested, generally accepted and used as the basis for the definition of other more complex standards. The second ones (emerging standards) are more recent, still under testing procedures and are not yet seen as the *de facto* solutions.

9.4 Existing Standards in grid computing

New and updated descriptions of standards from D4.3.1.

9.4.1 Extensible Markup Language (XML)

Defined by: W3C

Current version: 1.1

Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML (ISO 8879). It's a general-purpose markup language for creating special-purpose markup languages, capable of describing many different kinds of data. Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere.

9.4.2 Grid Security Infrastructure (GSI)

Defined by: IETF

Current version: -

The Grid Security Infrastructure (implemented by the Globus Toolkit) is a *de facto* standard for Grid security. GSI uses X.509 identity and proxy certificates, which provide a globally unique identifier that can authenticate and authorize an entity with accessed Grid resources. In GSI, the owner typically grants use of a resource to individual users, who must have an account for each accessed resource.

9.4.3 Lightweight Directory Access Protocol (LDAP)

Defined by: IETF

Current version: 3.0

LDAP provides access to distributed directory services that act in accordance with X.500 data and service models. This protocol is based on those described in the X.500 Directory Access Protocol (DAP).

9.4.4 Open Grid Services Architecture (OGSA)

Defined by: GGF

Current version: 1.0

OGSA is probably the most important Grid standard to emerge recently. It aims to define a common, standard, and open architecture for Grid-based applications.

OGSA defines what Grid services are, and the overall structure and services to be provided in Grid environments. Building on existing Web services standards OGSA defines a Grid service as a Web service that conforms to a particular set of conventions. For example, Grid services are defined in terms of standard WSDL with minor extensions. This is important because it gives us a common and open-standards-based set of techniques to access various Grid services using existing standards, such as SOAP, XML, and WSSecurity. With this base, can be added and integrated additional services (such as life cycle management) in a seamless manner. It provides a standard method to find, identify, and utilize new Grid services as they become available. And as an added benefit, OGSA will provide for interoperability between Grids that might have been built using different underlying tools.

9.4.5 Secure Socket Layer (SSL)

Defined by: IETF

Current version: 3.0

SSL is a security protocol that provides communications privacy over the Internet. The protocol allows client/server applications to communicate in a way that is designed to prevent eavesdropping, tampering, or message forgery.

9.4.6 Simple Object Access Protocol (SOAP)

Defined by: W3C

Current version: 1.2

SOAP is a lightweight XML-based protocol for exchange of information in a decentralized, distributed environment. SOAP can potentially be used in combination with a variety of protocols. However, the only bindings defined in this document describe how to use SOAP in combination with HTTP and HTTP Extension Framework.

9.4.7 Universal Description, Discovery and Integration (UDDI)

Defined by: OASIS

Current version: 3.0.2

The UDDI specification describes the Web Services, data structures and behaviours of all instances of a UDDI registry. It's a platform-independent, XML-based registry for businesses worldwide to list themselves on the Internet. UDDI enables businesses to publish service listings and discover each other and define how the services or software applications interact over the Internet.

9.4.8 Web Services Description Language (WSDL)

Defined by: W3C

Current version: 1.1

WSDL is an XML format for describing network services as a set of end points operating on messages containing document or procedure-oriented information. WSDL is extensible to allow description of endpoints and their messages regardless of what message formats or network protocols are used to communicate. However, the bindings in this document describe how to use WSDL in conjunction with SOAP 1.1, HTTP GET/POST, and MIME. Several drafts for version 2.0 are already available how should be approved by the W3C soon.

9.4.9 Web Services Interoperability Basic Profile (WSI BP)

Defined by: WS-I

Current version: 1.1

WS-I BP defines a set of non-proprietary Web services specifications, along with clarifications, refinements, interpretations and amplifications of those specifications which promote interoperability. It contains guidelines for using SOAP, WSDL, and UDDI. WS-I BP has both recommendations and requirements for compliant services (for example, it recommends sending SOAP messages with HTTP/1.1 but requires the use of either HTTP/1.1 or HTTP/1.0).

Many applications other than Web services use HTTP, which has features that are appropriate in some environments but not in others. For example, HTTP cookies facilitate Web-based state management, but because cookies are not part of the SOAP envelope, WS-I BP mandates their use only in limited ways.

In some cases, WS-I BP tightens requirements in existing specifications. For example, SOAP 1.1 allows the use of the HTTP POST method as well as the HTTP Extension Framework's M-POST method, whereas BP1.0 permits only the former.

BP also clarifies ambiguities in some specifications. For example, a service sends a SOAP fault message when an error occurs. BP requires that the soap:fault element has no element children other than faultcode, faultstring, faultactor, and detail. Further, for extensibility the detail element can contain any type of element, thus a compliant service must accept such messages.

9.4.10 Web Services Addressing (WS-Addressing)

Defined by: W3C

Current version: 1.0

Web Services Addressing provides transport-neutral mechanisms to address Web services and messages. WS-Addressing defines a set of abstract properties and an XML Infoset representation thereof to reference Web services and to facilitate end-to-end addressing of endpoints in messages. This specification enables messaging systems to support message transmission through networks that include processing nodes such as endpoint managers, firewalls, and gateways in a transport-neutral manner.

9.4.11 Web Services Management (WS-Management)

Defined by: DMTF

Current version: 1.0

WS-Management addresses the cost and complexity of IT management by providing a common way for systems to access and exchange management information across the entire IT infrastructure. By using Web services to manage IT systems, deployments that support WS-Management will enable IT managers to remotely access devices on their networks - everything from silicon components and handheld devices to PCs, servers and large-scale data centers. WS-Management is the first specification in support of the DMTF initiative to expose CIM resources via a set of Web services protocols.

9.4.12 Web Services Notification (WS-N)

Defined by: OASIS

Current version: –

This standard is still under the public review phase. It is expect to become as standard after passing this review.

These specifications standardise the way Web Services interact using "Notifications" or "Events". They form the foundation for Event Driven Architectures built using Web services. They provide a standardized way for a Web service, or other entity, to disseminate information to a set of other Web services, without having to have prior knowledge of these other Web Services. They can be thought of as defining "Publish/Subscribe for Web services". These specifications have many applications, for example in the arenas of system or device management, or in commercial applications such as electronic trading.

The Web Service Notification (WS-N) is composed by three specifications:

- WS-BaseNotification – handles asynchronous notification, including interfaces used by a notification producer or consumer
- WS-Topics – organizes and categorizes items of interest for subscription, known as topics
- WS-BrokeredNotification – handles asynchronous notification

The WSN specifications are not considered part of WSRF proper, but rather build on WSRF. The WS-BrokeredNotification is not currently supported by the Globus Toolkit but it is listed for completeness.

9.4.13 Web Services Reliability (WS-Reliability)

Defined by: OASIS

Current version: 1.1

WS-Reliability is a SOAP-based protocol for exchanging SOAP messages with guaranteed delivery, no duplicates, and guaranteed message ordering. WS-Reliability is defined as SOAP header extensions and is independent of the underlying protocol. This specification contains a binding to HTTP.

9.4.14 Web Services for Remote Portlets (WSRP)

Defined by: OASIS

Current version: 1.0

WSRP is a standard for Web portals to access and display portlets that are hosted on a remote server. This specification is a joint effort of two OASIS technical committees. Web Services for Interactive Applications (WSIA) and Web Services for Remote Portals (WSRP) aim to simplify the integration effort through a standard set of web service interfaces allowing integrating applications to quickly exploit new web services as they become available. The joint authoring of these interfaces by WSRP and WSIA allows maximum reuse of presentation-oriented, interactive web services while allowing the consuming applications to access a much richer set of standardized web services.

This joint standard layers on top of the existing web services stack, utilizing existing web services standards and will leverage emerging web service standards (such as security) as they become available. The interfaces are defined using the Web Services Description Language (WSDL).

9.4.15 Web Services Resource Framework (WSRF)

Defined by: OASIS

Current version: 1.2

WSRF is a set of proposed Web Services specifications that define a rendering of the WSResource approach in terms of specific message exchanges and related XML definitions. These specifications allow the programmer to declare and implement the association between a Web service and one or more stateful resources. They describe the means by which a view of the state of the resource is defined and associated with a Web services description, forming the overall type definition of a WS-Resource. They also describe how the state of a WS-Resource is made accessible through a Web service interface, and define related mechanisms concerned with WS-Resource grouping and addressing.

The WS-Resource Framework (WSRF) is composed by four specifications:

- WS-ResourceProperties – defines how to query and modify WS-Resources described by XML Resource Property documents
- WS-ResourceLifetime – describes how to manage the lifetime of a resource and specifies Web services operations used to destroy a WS-Resource
- WS-ServiceGroup – describes how to represent and manage collections of Web services and/or WS-Resources
- WS-BaseFaults – defines a base fault XML type for use when returning faults in a Web services message exchange

WSRF is a joint effort by the Grid and Web Services communities. It provides the stateful services that OGSA needs. While OGSA is the architecture, WSRF is the infrastructure on which that architecture is built on. A WSRF goal is to evolve the Grid architecture in a way that's more clearly aligned with the general evolution of Web services. Instead of defining a new type of Grid service, these specifications will allow the services specified in the OGSA to be based completely on standard Web services. [WSRFGT4]

9.4.16 Web Services Security (WS-Security)

Defined by: OASIS

Current version: 1.1

This specification describes enhancements to SOAP messaging to provide message integrity and confidentiality. The specified mechanisms can be used to accommodate a wide variety of security models and encryption technologies.

This specification also provides a general-purpose mechanism for associating security tokens with message content. No specific type of security token is required, the specification is designed to be extensible (i.e.. support multiple security token formats). For example, a client might provide one format for proof of identity and provide another format for proof that they have a particular business certification.

Additionally, this specification describes how to encode binary security tokens, a framework for XML-based tokens, and how to include opaque encrypted keys. It also includes extensibility mechanisms that can be used to further describe the characteristics of the tokens that are included with a message.

10 BIBLIOGRAPHY AND REFERENCES

- [1] C. Lagoze, S. Payette, E. Shin, and C. Wilper. *Fedora: An Architecture for Complex Objects and their Relationships*. Journal on Digital Libraries, Special Issue on Complex Objects, 2005.
- [2] DELOS Network of Excellence on Digital Libraries <http://www.delos.info>
- [3] Snowdon D. N., Churchill E. F., Frecon E. (eds) *Inhabited Information Spaces - Living with your Data*. Springer, London 2004.
- [4] S. Anderson and R. Heery. *Digital Repositories Review*. JISC Digital Repositories Programme, 2005. http://www.jisc.ac.uk/uploaded_documents/digital-repositories-review-2005.pdf
- [5] R. Tansley, M. Bass, D. Stuve, M. Branschofsky, D. Chudnov, G. McClellan, and MacKenzie Smith. *The DSpace Institutional Digital Repository System: Current Functionality*. In Proceedings of the 3rd Joint Conference on Digital Libraries, JCDL 2003, pages 87-97, IEEE Computer Society, 2003.
- [6] *Fedora Flexible Extensible Digital Object and Repository Architecture*. <http://www.fedora.info>.
- [7] aDORe J. Bekaert, P. Hochstenbach, and H. Van de Sompel. *Using MPEG-21 DIDL to Represent Complex Digital Objects in the Los Alamos National Laboratory Digital Library*. D-Lib Magazine, 9(11), November 2003.
- [8] L. Candela et. al, *Current Digital Library Systems: User Requirements vs Provided Functionality*, DELOS Deliverable D1.4.1, March 2006.
- [9] DELOS Reference Model Activity, http://www.delos.info/index.php?option=com_content&task=view&id=345
- [10] *OpenDLib A Digital Library Service System*. <http://www.opendlib.com>
- [11] C. Schuler, H. Schuldt, C. Türker, R. Weber, and H.-J. Schek. *Peer-To-Peer Execution of (Transactional) Processes*. In: International Journal on Cooperative Information Systems (IJCIS). Vol. 14, No. 4 (2005) 377-405.
- [12] M. Mlivoncic, C. Schuler, and C. Türker. *Hyperdatabase Infrastructure for Management and Search of Multimedia Collections*. In DELOS Workshop: Digital Library Architectures, pages 25–36, S. Margherita di Pula, Cagliari, Italy, 2004.
- [13] Daffodil. *Distributed Agents for User-Friendly Access of Digital Libraries*. <http://www.daffodil.de>
- [14] I. Foster, C. Kesselman, and S. Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations," *Int'l J. Supercomputer Applications*, vol. 15, no. 3, 2001, pp. 200-222.
- [15] B. Hughes. *Building Computational Grids with Apple's Xgrid Middleware*. Department of Computer Science and Software Engineering, University of Melbourne.
- [16] *Fedora Flexible Extensible Digital Object and Repository Architecture*. <http://www.fedora.info>.
- [17] Ioannidis, Y.: Digital libraries at a crossroads. *International Journal on Digital Libraries* 5(4), 255–265 (2005)
- [18] T. Staples, R. Wayland, and S. Payette. The Fedora Project: An Open-source Digital Object Repository Management System. *DLib Magazine*, 9 (4), April 2003.
- [19] R. Tansley, M. Bass, and MacKenzie Smith. *DSpace as an Open Archival Information System: Current Status and Future Directions*. In Proceedings of the 7th European Conference on Research and Advanced Technology for Digital Libraries, ECDL 2003, pages 17-22, Trondheim, Norway, August 2003.
- [20] R. Tansley, M. Bass, D. Stuve, M. Branschofsky, D. Chudnov, G. McClellan, and MacKenzie Smith. *The DSpace Institutional Digital Repository System: Current Functionality*. In Proceedings of the 3rd Joint Conference on Digital Libraries, JCDL 2003, pages 87-97, IEEE Computer Society, 2003.

