SIXTH FRAMEWORK PROGRAMME

PRIORITY IST-2002-2.3.1.12

Technology-enhanced Learning and Access to Cultural

Heritage

Contract for:

NETWORK OF EXCELLENCE

Annex 1 - "Description of Work"

Network acronym: **DELOS**

Network full title: DELOS: a Network of Excellence on Digital Libraries

Proposal/Contract no.: **G038-507618**

Related to other Contract no.: IST-1999-12262

Date of preparation of Annex 1: **01 December 2003 (FINAL)**

Operative commencement date of contract: **01 January 2004**

1   Outline of the JPA for the full duration of the Network

The Joint Programme of Activities is organized into the following seven clusters of integration and research activities, which correspond to seven Work Packages. The dissemination activities (Spreading of Excellence) and the management activities have been grouped in separate clusters (Work Packages), described in the subsequent sections.

Short name	Cluster name	Cluster Coordinators
ARCH	Digital Library Architecture	Prof. Hans-J. Schek (ETH and UMIT)
IAP	Information Access and Personalization	Prof. Yannis Ioannidis (UOA)
A/V-NTO	Audio/Visual and Non-traditional Objects	Prof. Alberto Del Bimbo (UniFI) Prof. Stavros Christodoulakis (TUC)
UIV	User Interfaces and Visualization	Prof. Tiziana Catarci (ROMA1)
KESI	Knowledge Extraction and Semantic Interoperability	Prof. Liz Lyon (UKOLN)
PRESERV	Preservation	Prof. Seamus Ross (UG)
EVAL	Evaluation	Prof. Norbert Fuhr (UNIDU)

A CV of the cluster coordinators is in Appendix A1.1. In the following subsections we describe the integration and joint research activities planned for each research cluster.

1.1   Integrating Activities

1.1.1   Digital Library Architectures

A core requirement for digital libraries is a common infrastructure. From a technical viewpoint, this infrastructure has to support state-of-the-art and promising innovative models and techniques, and frameworks to develop and evaluate digital libraries, and has to be highly customizable, configurable and adaptive. To this end, various activities and developments have to be seamlessly integrated into a coherent whole to develop such a generic and modular digital library infrastructure. This includes the following **architectural approaches, processes, and activities**.

Architectural Approaches
Peer-to-Peer Architectures: Information providers within digital libraries are highly autonomous. Data and documents cannot be integrated into a single source. Hence, mechanisms to retain this autonomy and to loosely couple information providers are needed. This also includes the user clients so as to facilitate some collaborative data sharing among them (e.g., for annotations and recommendations about DL contents). Peer-to-peer (P2P) architectures allow for such loosely coupled integration. Different aspects of peer-to-peer systems (e.g. indexes, and P2P application platforms) have to be combined and integrated into an infrastructure for digital libraries.

Grid Architectures: Certain services within digital libraries are complex and computationally intensive (e.g., calculation of certain features of multimedia documents to support content-based similarity search). Grid computing architectures allow for sophisticated load balancing strategies within a cluster of components. Following the idea of a service grid, and the handling of the control of shared resources, similar concepts have to be integrated into an infrastructure for digital libraries.

Service-oriented Architectures: When access to data and documents is provided by dedicated services, appropriate mechanisms to describe the semantics and usage of such services have to be put in place. In the context of web services, descriptions of services using service description languages, are stored in service registries. These elements have to be integrated as building blocks into a digital library. Moreover, common service interfaces have to be defined based on existing standards to facilitate service composition.

Processes

Workflow Management: Applications within digital libraries must consider the autonomy and distribution of information providers. Hence, accessing information means combining existing services into mega-applications, i.e. workflow processes. The same is true for applications aiming at managing and controlling the consistency of a digital library. Different aspects of workflow management have to be integrated: such as self-configuration and flexibility, both at the application and at the systems level, and high availability and scalability.

Publish/Subscribe Techniques, Evolution: Services within a digital library have to be made available to the public and have to be accessed by service repositories. Publish/subscribe techniques are a means to make information within digital libraries available and to refresh derived and replicated information sources. Digital libraries are long-lasting institutions, so they have to anticipate changes in the software as well as in the schemata, the ontologies and similar data. The infrastructure for digital libraries therefore has to provide mechanisms to distribute and co-ordinate updates to these components, and to manage the software configuration in such a dynamic environment.

Replication and Freshness of Data: In order to increase the efficiency in accessing information within a digital library, information will be replicated at several places. There will also be duplicates due to independent upload of the same publication at different nodes. However, when changes occur (new data is provided or information is updated), this has to be reflected in all replicas, leading to a trade-off between update costs and freshness of data. Sophisticated mechanisms to trade, in an application-specific way, update costs for the freshness of data need to be provided by the digital library infrastructure.

Mobile Information Components: The combination of wireless and wired connectivity in a pervasive computing environment with increasingly small and powerful mobile devices, such as laptops, personal digital assistants, handheld PCs, and smart phones, enables a wide range of new digital library applications. This additional flexibility has to be supported by the underlying digital library middleware. In particular, profiling and proxy management have to be integrated into the infrastructure. Additionally, mobile devices will require sophisticated visualisation techniques to present digital publications adequately on limited displays.

Functions

Data and documents within a digital library are made available by dedicated services. These services allow for the definition of building blocks, which are tailored to the type of data and documents and implement, for instance, appropriate index structures.

XML Storage and Access: Effective and efficient access methods for documents in XML stores will provide the basis for mediation within P2P information architectures. The emerging language for annotation of digital content is XML. Its power to annotate any document challenges the techniques to store, index and access them. Progress is required in areas such as IR techniques over XML sources, shredding and selective indexing for fast retrieval, clustering, replication management, and transformation aimed at improved transport and platform specific delivery.

Multimedia Access: A key task of a digital library is the maintenance and retrieval of documents of various types and for different search scenarios. Due to the distributed nature, services from

different nodes must be able to interact with each other. To this end, well-designed service interfaces are required to ease integration of different providers, e.g., of feature extraction algorithms, indexing services, and retrieval engines. Note: techniques to extract features, and maintain and index such features are covered by other clusters.

Digital Rights Management: Publishers will be reluctant to provide content if there is no digital rights policy enforceable. This includes support for business models such as pay-per-view or subscriptions. Peer-to-Peer (P2P) architectures have a notoriously bad reputation in this respect from their application in the music industry context and this situation needs to be overcome to facilitate business development. Replication within digital libraries adds to the complexity of this aspect.

Security and Certification: The use of Digital Rights Management in autonomous environments such as digital libraries based on peer-to-peer architectures requires the authentication and authorization of users, authors, content providers, and reviewers as well as digital signatures for documents, to ensure the consistency, quality, and reliability of digital libraries.

Finally, the results of the different activities of the architecture cluster need to be integrated and evaluated in concrete settings. To this end, demonstrator systems and building blocks have to be combined. The cluster will check all the solutions that will be developed in a sample application domain, that is currently being analysed by some of the cluster members and that appears to be well suited to evaluate different digital library platforms and infrastructures.

Medical Information Systems: E-Health Warehouses can be considered as special digital libraries. Patient-related information is available from different distributed and autonomous information providers. In order to provide electronic patient records, this information must be integrated at application level. Several issues addressed in the joint research activities are important: appropriate service interfaces of building blocks, workflow management for application development, P2P infrastructures, replication and freshness, security and certification, etc.

1.1.2   Information Access and Personalization

Information stored in digital libraries needs to be accessed, integrated and individualized for any user anytime and anywhere in possibly multiple comprehensive and efficient ways. Consequently, Information Access in Digital Libraries may be studied from three different aspects: (i) Access of information stored in an individual source, (ii) Integrated Access of information distribution in different sources and (iii) Access of information for different users. Within the Information Access and Personalization cluster (abbreviated as IAP cluster), integration activities will primarily focus on collection, study, comparison, promotion and integration of appropriate models, (representation or query) languages and algorithms for supporting the above-mentioned issues. The main objectives for these activities are the following:

- Promotion of knowledge about available practices in the fields of information access and personalization in digital libraries.

- Provision of a common foundation for different digital library access approaches that will lead to a uniform understanding of problems among researchers and stimulation of joint research efforts.

- Support of cooperation among individual research groups that will promote common consensus among researchers.

Common Foundation on Information Access

Information stored in a source comes in different types and formats, e.g., unstructured vs. semi-structured, business vs. scientific etc, each one with its own characteristics and particularities. Organization of data within an individual source and efficient and effective search are the key issues and are actually highly interrelated to each other. Different approaches exist but there is a general trend towards richer representations and languages both at the structural and at the semantic level. In this context, the role of metadata, which may again come into different representations, is very important. The IAP cluster plans integration activities that aim at the establishment of a common foundation on Access of different types and formats of Information, specifically focusing on the following research problems:

Data and Query Models and Query Processing Schemes: Efficient data modeling and appropriate query languages and query processing techniques will be investigated by the IAP cluster with respect to different data forms, i.e., structured, semi-structured and unstructured. More specifically, given the affluence of existing approaches and tools, the cluster will work towards a comprehensive collection and categorization of them. More importantly, advantages and drawbacks will be highlighted and potentially new or improved methods and models will be identified, some of which will be further developed under specific joint research activities. XML, which is now standard for semi-structured data and metadata, will play a central role in these activities.

Metadata Models and Metadata Query Processing Schemes: Metadata capture additional knowledge about a digital library. Current trends deal with the use of different types of metadata in order to enhance and improve information access. Taxonomies, ontologies, thesauri as well as other abstraction mechanisms represent different kinds of additional knowledge a source may maintain and will be studied and compared by the IAP cluster. More specifically, existing metadata models and metadata query schemes will be studied with the intention to create a classification of them with respect to the role and usability of each one of them in a DL setting as well as to identify advantages and shortcomings of each approach. (These activities will take place in co-operation with the Audio/Visual cluster, as well as the Knowledge Extraction and Semantic Interoperability cluster).

Metadata Manipulation: As metadata become very important in a DL environment, definition of efficient and appropriate manipulation operations is essential. This cluster will study existing research in this field and work towards the specification of a set of logical operations that can be meaningful and useful for metadata manipulation. (These activities will take place in co-operation with the Audio/Visual cluster for multimedia metadata).

Data and Metadata Generation: In the context of a digital library, creation and maintenance of data as well as metadata may be supported manually or (semi-) automatically. Metadata may be generated out of data or even existing metadata. The IAP cluster will perform a concise study of non-manual algorithms and methods, such as crawling, clustering and mining and record the potential as well as the role of each of them in a DL setting. A list of possible improvements or directions of research may be produced. (These activities will take place in co-operation with the Audio/Visual cluster for multimedia metadata as well as the Knowledge Extraction and Semantic Interoperability cluster).

Specific Data Types: Specific data types such as scientific data, in particular sensor data streams, may represent different research challenges from other conventional (commercial/administrative) DLs. The IAP cluster will study requirements and special characteristics of certain types of data with focus on scientific data and provide a list of possible issues requiring further investigation by the research community.

Common Foundation on Information Integration
Integrated access of different sources presents specialized problems. Different approaches and scenarios exist. For example, one may consider each source having a representation of the semantics of its contents that is a fragment of a global scheme. On the other hand, other approaches exist that are not based on such global assumption. The IAP cluster plans integration activities (that will take place in co-operation with the Architecture cluster) aiming at the

establishment of a common foundation on Integration of Information coming from more than one digital library, specifically focusing on the following research problems:

Integrated Query Processing Schemes: Query evaluation over distributed information sources presents particular problems due to information heterogeneity, redundancy etc. Issues such as source selection and results fusion must be considered under different possible settings as described above. The IAP cluster will study and classify available approaches with respect to advantages, shortcomings, and the different settings under which they may be efficient and useful. As a direct result, possible improvements in algorithms as well as general directions of future research will be highlighted. (These activities will take place in co-operation with the Audio/Visual cluster for multimedia metadata).

Integrated Metadata: In order to provide integrated access over distributed Digital Libraries, representation of metadata over a set of sources may be necessary, in the form of an ontology, a thematic hierarchy etc. The IAP cluster will study and consider all issues previously studied in the context of a single source (i.e., models, manipulation operations, acquisition methods) but focusing on the particularities stemming from the distribution and heterogeneity of information. (These activities will take place in co-operation with the Audio/Visual cluster for multimedia metadata).

Data Provenance: Assembling and publishing data may be now cheap and easy. The problem is that as data gets copied, transformed and edited one loses track of where it comes from. This information, termed data provenance, is often crucial to the trust that is placed in the data, and may be useful in scientific discovery. Yet there are no tools for tracking provenance on the Web, and databases, being rather rigidly structured, typically do not provide room to accommodate provenance information. The IAP cluster will describe and study the problem and existing approaches and will provide directions for further research in the field. (These activities will take place in co-operation with the Knowledge Extraction and Semantic Interoperability cluster)

Common Foundation on Personalization
Different users have different characteristics and preferences concerning the information they are interested in seeing when accessing a digital library. Even users sharing a common information need may expect different results, different functionality or different interface. Moreover, the relevant contents and interface of a digital library may be dependent on other factors as well, e.g. device or network-specific. The IAP cluster plans integration activities with respect to the following research problems on the user aspects of information access:

User Characteristics for Personalization: A thorough and broad study of the relevant user characteristics (such as user behavior, preferences, location etc) is essential and is missing, though there is quite a lot of experience from the deployment of many digital libraries systems. The IAP cluster will work towards the collection of information concerning user characteristics regarding various aspects of a digital library that affect the entire user experience; e.g., information organization, presentation, retrieval and personalization functionality. Such an activity will open the path for the specification of a conceptual model of personal digital libraries that will be used as reference point of any effort in this direction. (These activities will take place in co-operation with the Audio/Visual cluster for consideration of multimedia).

User Modeling for Personalization: In order to provide access to a digital library setting tailored to the needs and preferences of a single user, information about the user must be recorded in profiles. Different models have been proposed so far in the literature. Moreover, different methods exist for creation and maintenance of machine-processable profiles. The IAP cluster plans a concise study of them. The final output of the effort will be a high-level specification of more general user models that are needed in digital libraries. (These activities will take place in co-operation with the Audio/Visual cluster for consideration of multimedia).

Content and Interaction Personalization: Information access may be tailored to a specific user or group of users at two levels: at the content level and at the interaction level. Annotations added by a user on documents and information filtering based on user preferences (i.e., what a user is

interested in) or privacy requirements (i.e., what a user is allowed to see) can be seen as different types of content personalization. Different functionality and personalization schemes are viewed as forms of personalization at the interaction level. The IAP cluster will record and characterize all possible forms of content and interaction personalization and highlight issues and areas that need further investigation sketching possible solutions wherever possible. (Rendering of information to the user is not the focus of this cluster, for that reason cooperation with the User Interfacesand Visualization cluster will be necessary for this part of personalization.)

Complementary to and as a result of the above activities, the following actions are planned:
Creation of a shareable common IAP Knowledgebase: The results and studies of the abovementioned activities will be made accessible in different forms, e.g. as written reports, or on the website and will be created either through the establishment of focused working groups and/or thematic workshops of invited specialists. Exchange of researchers and PhD students may also take place in several cases.

IAP Community support: The IAP cluster will support activities for stimulation and facilitation of discussion and share of knowledge among research groups. These activities may include construction of FAQs, dedicated message boards, periodicals (electronic and/or hard copy), organization of summer schools.

Integrated Prototypes: The IAP cluster will set up a set of prototype systems among two or more research groups that will combine existing expertise and tools. The aim of this effort will be twofold: (a) to help different groups make use of and build on others' expertise and developments (b) to combine different approaches for common problems and measure possible improvements. Useful results from these prototypes will be communicated to the rest of the community. A close cooperation with the architecture cluster and the prototype development there is planned. The integration of various access modules and services into the open architecture serves both clusters as an important test case for building digital libraries.

1.1.3   Audio/Visual and Non-traditional Objects.

Digital libraries will organize, store and manage large amounts of human knowledge in different application domains, for various uses and user communities, providing access to their content in various contexts, from a variety of delivery channels.

Digital libraries will often store non-traditional content, such as images, sound, video, graphics, etc. The non-traditional content will often be highly structured into segments and/or semantic units (objects) which will be indexed and interconnected with other objects in a variety of ways allowing flexible access, transcoding, browsing, semantic integration, presentation, and personalization according to the application functionality, the domain of knowledge described, the presentation device and the user's preferences and goals. Application areas of digital libraries with non-traditional content include art, history, culture, medicine, natural and geo-sciences, engineering, public sector, etc They are bound to have significant impact in the knowledge society

Digital libraries require appropriate definition and management of their lifecycle. The lifecycle of "content" of a digital library includes the following major stages of: 1) Ingestion/creation (capturing physical objects into digital representation), 2) Editing – (normalization, standardization and cleansing of the captured data), 3) Analysis (metadata extraction and their correlation with other "related" content), 4) Management (of both metadata and content, including the development of indices for faster retrieval, data integrity, consistency, and versioning), 5) Distribution (content dissemination, copyright management, etc.). Lifecycle management will therefore include capturing and tracking of the workflow associated with all aspects of content so that the processing steps of the content from creation to dissemination can be better automated with richer, more intelligible, results. Users may want to access non-traditional content from heterogeneous distributed digital libraries, and compare and synthesize the accessed information. The access to information will be done through a variety of delivery channels and the

information delivered will often be associated with physical objects taking into account the context.

The scope of the work that will be done in the area of audiovisual content in digital libraries will be focusing on metadata capturing for audio-visual content, universal access and interactions with audiovisual libraries, and management of audio-visual content in digital libraries.

Common Foundation on Metadata Capturing for Audio-Visual Content
To manage audio-visual content effectively one has to extract and organize metadata related to the content. The metadata extracted have to be based on models supported by the developing audio-visual content description frameworks (like MPEG-7 and TV-Anytime) and they have to take advantage of domain and context specific knowledge. The metadata extraction process also has to take advantage of such domain and context specific knowledge. The following topics will be addressed by the audio-visual cluster:

Automatic Multimedia Information Extraction: The cluster will follow developments in the state of the art of automatic information extraction algorithms and tools for multimedia content. . (Some of these activities will take place in co-operation with the Information Access and Personalization cluster).

Domain and Context Specific Information Extraction: Domain-specific and context-specific methodologies to achieve more automation and/or higher performance plus better description quality in the extraction of content describing metadata will be traced by the cluster.

Models of Multimedia Content and Context: The cluster will survey and unify the developing state of the art in models of multimedia content and context, and the standards for the description of audiovisual content and ontologies. The state of the art of research and results in the fields of databases and information retrieval will be studied with respect to metadata for audiovisual content, to achieve effective retrieval in digital libraries.

Common Foundation on Universal Access and Interactions with Audio-Visual Libraries
The delivery of digital library information through diverse channels like the web, TV (broadcasting) and mobile devices (wireless) and the use of digital library information in different locations and contexts will have a major impact in Digital Library adoption. The audio-visual cluster will study the developing state of the art in the use of audio-visual digital library content in the diverse delivery channels, using diverse interaction devices.

Multimodal Interfaces to Multimedia Content: Paradigms, solutions and methodologies to interface with multimedia content in diverse delivery channels using diverse devices will be surveyed.

Interaction with Multimedia Content: Users can interact with multimedia content using diverse and sometimes delivery channel dependent devices from voice to TV control or eye sensing. The cluster will survey developing technologies for delivery and presentation of multimedia data over different channels and divvent devices. The cluster will also study cognitive models explaining user behavior and intentions. (These activities will take place in co-operation with the User Interface and Visualization cluster)

Delivery of Multimedia Content: Models of delivery and solutions (functionality of platforms and algorithms) for the delivery of multimedia content across diverse platforms including web, peer to peer, digital TV and mobile will be surveyed.

Common Foundation on Management of Audio-Visual Content in Digital Libraries
Multimedia Content Description Standards and Domai-specific Extensions: The cluster will survey the multimedia content standards and the domain-specific extensions proposed, as well as the mapping efforts for providing interoperability across standards, domains, and applications.

Multimedia Summarization and Dissemination: Intelligent summarization and segmentation of multimedia content, compatible to standards using domain specific knowledge will be surveyed. Multimedia domain aware user profiling and stereotyping methodologies will also be surveyed.

Since the multimedia content description and the user profiling follow the multimedia standard frameworks (MPEG-7, TV-Anytime, etc), design methodologies for structuring of metadata for efficient matching and retrieval can accelerate the retrieval performance. The cluster will survey such design strategies.

Communities of Users of Digital Libraries: Communities having common interests in thematic subjects are important for knowledge organization, knowledge acquisition and community based filtering of information. The cluster will survey the state of the art in these fields. (These activities will take place in co-operation with the Information Access and Personalization cluster as well as the User Interfaces and Visualization cluster).

1.1.4   User Interfaces and Visualization

The notion of a "Digital Library" (DL) is currently associated with technological and scientific efforts to build, maintain, and use large collections of electronic documents. However, it can also be regarded as a cornerstone in the construction of an information-enriched environment. Once this broader perspective is adopted, a variety of problems arise which will have to be solved in order to ensure the usability of this environment for professional and recreational purposes (cf. Duguid & Atkins 1997). Accordingly, central DL research issues include, for example, human-centered research (retrieval models, intelligent interfaces, collaboration), and content and collections-based research (e.g. metadata, preservation, domain-based content analysis), in addition to systems-centred research (e.g. network architectures, interoperability) [cf. U.S. Digital Libraries Initiative - Phase 2]. For instance, the DL will be accessible to different users with varying needs and backgrounds, and provide access to heterogeneous collections containing items of probably very diverse contents, type, validity and genre. The design and the functionality of the user interfaces of the DL ought to reflect this diversity in order to provide information in the right form at the appropriate time and place. The adopted perspective stresses the fact that "information" is not just static data or documents but the outcome of a context-dependent communication process involving humans, documents, and the technological as well as informational environment. Thus, the way a DL operates - or should operate - is heavily dependent on factors which constitute new requirements for the design of information systems. The proposed NoE aims at enabling future DL designers and developers to meet not only the technological, but also the user-oriented requirements in a balanced way. The cluster will address this goal by pursuing several integration activities:

User requirements
Provision of an empirical basis: Within the proposed NoE, user requirements will be studied systematically. The different perspectives on a digital library - ranging from a traditional point of view that focuses on collections of information items made accessible by query forms and browsing facilities to the notion of a virtual environment that provides a spatio-temporal arrangement of the items visualized using appropriate metaphors (virtual buildings, rooms etc, or more abstract topic maps) - will be analysed to relate them to the requirements and technical implementation options that emerge from the ongoing development projects of the NoE partners participating in the User Interfaces and Visualization cluster and in the other NoE clusters. This analysis will be complemented by: state-of-the-art reviews which result from DL literature, workshops, and task forces. It will also include input contributed by users from relevant organizations (e.g. museums, schools, archives, virtual organizations, industries); and input from related research efforts/projects.

Support for All Phases of DL Lifecycle (A homogeneous all-phase encompassing environment): Future DL solutions will be required to support all the phases pertaining to the entire DL lifecycle. This activity will analyse all the aspects and phases pertaining to the development and the usage of a DL system. The analysis will not focus only on the DL end user but will also take into account other DL stakeholders such as librarians, content providers and maintainers. The activity will investigate DL lifecycle phases and aspects such as system functionality; user goals; tasks and activities such as scanning, indexing, thesaurus support, multimodal support, multilingual support,

searching and browsing, multilevel user help, integration of views/results from various libraries and library search tools; and non-functional aspects such as platform independence and scalability. In cooperation with the Information Access and Personalization cluster, user adaptation during system usage too will be explored. As far as the multilevel user help is concerned, it would be interesting to consider a "Virtual Users' Guide Library" in the NoE by cooperating with the Virtual Digital Library Competence Centre.

Characterization of DL users (e.g. librarians, end users) and stakeholders (e.g. content providers, maintainers, funders) and their diverse needs: It is crucial that these requirements be taken into account throughout all the phases of the DL lifecycle. So far, most of the work in the DL domain has concentrated on addressing the needs of the end user with little, if any, consideration for other DL stakeholders (such as librarians, archivists, and content providers). However, the role and the experience of such stakeholders are of great importance and must not be neglected; they represent a class of users with different needs from that of the end users. Moreover, there is a need to guarantee universal usability of DLs by ensuring that the user interface accords accessibility for all categories of users, including users with special needs (e.g. users with different types of impairments/disabilities and the elderly). In addition, and coherently with the grand vision of DELOS for DLs, this cluster will also explore how users can exploit a multi-modal DL user interface to meet their particular needs.

We recognize that users as individuals are unique and hence each has his/her own specific needs. However, different user communities also have different needs. They require different levels/abstractions of information and different user interface interactivity models. The user interface should be in a position to capture user's/community's context, maintain user's/community's profiles and facilitate the exchange of information so as to enrich and assist users/communities during later DL tasks. This cluster will thus work closely with the other clusters. It will take the user and/or interaction models defined by the other clusters and interpret them for the DL user interface.

User Interface and Visualization Design
Context Consideration and Exploitation: The user's environment (physical, organizational, social, technical) presents a resource that may be exploited to meet his/her needs. Developments in telecommunications, and in particular the Internet, mobile devices and wireless technologies, have opened up incredible frontiers for DLs. One consequence is that DLs are now dealing with a wider user audience and an increasing number of user communities. These users/communities would appreciate capabilities enabling them to exploit knowledge based on collaborations with other DL users/communities. The DL user interface should thus support the user in collaborative work. Another implication of the technological advances is that space and distance are no longer the barriers they used to be. The implication is that we need interfaces that can support the user to access the digital library from anywhere, at anytime, and while s/he is carrying out some other activity. In a more general sense, we will explore the impact and corresponding DL opportunities with respect to contextual information. The corresponding activities will take place in co-operation with the Audio/Visual cluster for multimedia metadata, as well as the Information Access and Personalization cluster.

Visualization: DLs store a vast amount of information. Moreover, complex relationships often exist among data items, which can involve complex types such as sound, video stream files, etc. It is therefore no surprise that a typical query to a DL may yield an overwhelming set of results in terms of quantity and complexity. Efforts from information visualization have helped to interpret complex and even massive sets of data. By definition, information visualization is the process of transforming data into a visual form that will assist or trigger the human user's visual capabilities to gain insight and understanding of that data. Mapping between the data and the visualized elements should ensure that conceptually important aspects become perceptually important. Moreover, the mapping should be done in a precise, consistent way. DLs have not yet exploited fully the results already achieved by information visualization (e.g TileBars by Marti Hearst). This cluster will investigate the exploitation of relevant existing and/or novel visualizations in presenting DL results/views, and even various aspects of the DL lifecycle. The possibility to

effectively extend the visualizations by using text and multimedia data will also be considered. The effectiveness, expressiveness and interactivity of visualization in the DL context will be major issues. These activities will be carried out in cooperation with the Knowledge Extraction and Semantic Interoperability cluster, and the Evaluation cluster.

Theoretical Framework for DL User Interface Design: This cluster will build a theoretical framework from which user interface designers/developers can design DL user interfaces. The designer/developer gathers various resources provided by the theoretical framework (e.g. methodologies and tools) and designs a DL user interface (e.g. tailored for some particular application domain).

1.1.5   Knowledge Extraction & Semantic Interoperability

Digital libraries contain different types of resources: text, multimedia, data, learning objects, and serve many users with different needs, e.g., a digital library for cultural heritage addresses researchers, students, visitors, school children, and the media. Each group has its own needs and requires different levels (abstractions) of information (which may be personalized) and different models for interaction with the system. It is essential that processes and mechanisms are in place to allow creators/producers/users/consumers to deposit materials into a digital library, access existing materials effectively and use and re-use resources in a variety of ways. There are a number of research themes which aim to maximize both the availability and the re-usability of resources and to build on the body of knowledge available to all. The development of repositories of information is gaining momentum at both national and institutional levels and software and systems designed to manage these resources are being widely implemented.

The role of this cluster will be to co-ordinate a range of activities in this topic area, to bring together research expertise and to facilitate the sharing of experience and understanding of the challenges in this arena. In addition, there are a number of activities which interface closely with tasks in other clusters, leading to increased integration and cohesion within the Network.

Open Access to e-Print Repositories: There has been a concerted effort to promote open access to the research literature with the success of the Open Archives Initiative, the latest release of the ePrints software from the University of Southampton, UK, the establishment of the European-focused Open Archives Forum and initiatives such as DARE. There has also been a drive to promote institutional repositories as the location for e-print deposit e.g. University of Cambridge DSpace project. These developments have all been made possible through the implementation of the OAI-Protocol for Metadata Harvesting within the information architectures. This activity will study the developments needed for progress in this area.

Open Access to Data Repositories: The primary focus to date in this area has been on e-prints and other textual materials such as theses, however, digital resources published in this way may include primary research data, experimental Grid data, gene and protein structure data, statistical data, satellite data, census data, environmental modelling data etc. The current increase in Grid-enabled applications is resulting in large volumes of data being collected in data libraries and this trend is likely to continue in the future. As a result, new scientific disciplines are evolving such as Astroinformatics, which are developing around research which is data-centric and predictive, in contrast to the observational, descriptive science of the past. These large datasets need to be managed, curated and made accessible to the research community. This activity will study the application of the OAI approach to data repositories.

Open Access to Learning Object Repositories: In parallel to the development of repositories of research data and derived information, many institutions are creating learning objects for manipulation and inclusion in learning programmes and curriculum-based activities. Learning Management Systems are being deployed as vehicles for the development and distribution of online courses as part of e-learning initiatives. Repositories of learning objects are developed both at national and institutional level, to enable the access to and deposit of discrete learning objects

for wider use by the community. This activity will study the application of the OAI approach to learning object repositories

Provenance: The integrity, authenticity and value of the mass of information and knowledge derived from original data, is actually dependent on a number of critical factors. For example in science, the provenance or origin of a particular set of data is essential to determining the likely accuracy, currency and validity of derived information and any assumptions, hypotheses or further work based on that information. Significant research has been carried out on describing the provenance of scientific data in molecular genetics databases, and the topic has recently been explored in the Global Grid Forum (GGF6) in relation to Grid data. The Open Archives Initiative has carried out work to describe the provenance of harvested metadata records and the concept is included as an element in the administrative metadata which is part of the METS metadata standard. The critical factors include the definition and acceptance of appropriate frameworks for metadata description, a shared understanding of the concept of provenance, the widespread use of unique identifiers, appropriate linking technology and the application of common ontologies for discrete domains. These concepts are relatively new but have the potential for significant impact on the way in which research and learning is conducted in the future and on the ability to integrate and re-use digital resources in a variety of ways. This activity will be integrated with an activity of the Information Access cluster that studies data provenance from the viewpoint of provenance tracking.

Semantic Web, Ontologies and Metadata Schema Registries: In order to achieve semantic interoperability between descriptions of services, collections and items, there needs to be a shared understanding of the meanings of terms and descriptors. Frequently, discrete subject domains have their own shared vocabulary, however specific terms may have different meanings within another subject domain. This challenge will be addressed in a number of ways, through the formation of sets of high-level subject terms and thesauri, through the creation of metadata schema registries where details of a domain schema can be published into the community, and through the creation of mapping tools which map elements of one schema onto those of another schema to facilitate interoperability. . (These activities will take place in co-operation with the Audio/Visual cluster for multimedia metadata, as well as the Information Access and Personalization cluster).

Knowledge Extraction: The increasing richness of both data and the descriptive metadata contained in digital libraries offers great potential for the application of a variety of tools to extract additional information to contribute to knowledge. The research community has a growing requirement for data manipulation tools to facilitate spatial change (federation, aggregation, dis-aggregation, replication, manipulation, linking, annotation, editing/versioning, transformation) and for knowledge extraction which can include analysis (textual, musical, statistical, mathematical, visual, chemical, gene), mining (text, data, structures), and modelling (economic, mathematical, biological). Taking an example, text mining techniques have been applied to resources in various domains and in particular to biomedical materials. Similarly, data mining techniques have been applied to domain datasets such as biomedical and physical data and this form of analysis is becoming increasingly important in the understanding of outputs from Grid-enabled projects and associated data repositories. This activity will focus on the development of tools to extract additional information from DLs to contribute to knowledge. (These activities will take place in co-operation with the Audio/Visual cluster for multimedia information, as well as the Information Access and Personalization cluster).

1.1.6   Preservation

Over the past eighteen months ERPANET, an Accompanying Measure funded under FP5, has brought together experts to investigate core issues in digital preservation and conduct studies to help it define best practices. What is evident is that not only is there a fragmentation of knowledge and skills amongst the stakeholder communities about how to handle existing digital preservation problems, but there is a fragmentation of research. New efforts to synergize, focus,

and promote digital preservation research are essential. Digital Libraries depend on preservation of the digital materials they contain and the ability to build successful digital libraries depends upon methodological and technical solutions. Beginning in 2002 an international workgroup brought together by DELOS and the National Science Foundation defined a research agenda for Digital Preservation and Archiving in broad terms. The DELOS Preservation Cluster will thus aim to:

- (inter)connect people (researchers, stakeholders, suppliers, vendors etc.), organisations and projects that can deliver this research agenda;

- co-ordinate and promote research and projects;

- enable identification, collection, and dissemination of information, knowledge and expertise;

- provide mechanisms for developing testbeds and metrics for assessing the effectiveness of preservation strategies, and tools for evaluating digital preservation strategies;

- create a coherent platform for proactive co-operation, collaboration, exchange and dissemination of research results and experience in the preservation of digital objects;

- relate the research agenda more directly to the development of exploitable product opportunities and develop links with the industrial sectors;

- eliminate the duplication of effort between the various research activities by creating an integrating framework; and

- ensure that the work of the digital preservation cluster has a direct impact on digital library architecture and development work of other clusters in the DELOS network.

Achieving these goals will be difficult and the Preservation Cluster will work towards this objective through a series of focused collaborative activities.

Digital Preservation Testbed Forum. Digital preservation research has been, with some minor exceptions such as NEDLIB, small scale and lacked the benefit of a testbed environment to foster the creation of metrics to measure the effectiveness of different preservation strategies in different digital library contexts. File formats provide an important example of an area where we need testbeds. As well as producing evidence as to the properties of files or data types that enable or put preservation at risk, we need predictive measures to enable developers to assess the preservation impact of attributes of formats in advance of their completed development. This work can only be carried out in a testbed environment. The problem is not just a lack of testbeds, but the lack of agreed metrics for establishing a distributed testbed environment. The work on digital repositories will depend on the establishing of environments where different solutions can be evaluated and the testbed will provide a crucial mechanism to move digital preservation technologies from the research arena to the services domain. The workgroup aims to produce, test and disseminate testbed design, validation, comparability metrics and a test data set. This activity will be conducted in collaboration with the evaluation cluster.

File Format Data and Typology. Studies recognise that insufficient research has been directed at determining the properties of formats for representing digital entities that pose a greater or lesser challenge to their preservation. Different formats require different kinds of strategic approaches to ensure that they can be accessed in the future. Problems with formats are exacerbated by the fact that archival collections that need to be managed as a whole generally contain entities in multiple formats; these formats have different rates of obsolescence. Easily accessible sources of information about digital file formats (e.g. syntactical and semantic form) have not been the subject of systematic, consistent, comprehensive, and quality assured data collection practices. Long term access to data formats must be 'completely documented and publicly available'[1] if we

---

[1] Peter Keller-Marxer (Swiss Federal Archive)

are to ensure that we can provide mechanisms to render digital information in the future. There is interest in the US through the Digital Library Federation, CLIR, and the Library of Congress in creating mechanisms for collecting information about file formats. There is a relationship between file formats with particular attributes and preservation strategies. For digital libraries to exploit this information we need to know much more about classes of file formats so we can relate these to preservation approaches. This activity will contribute to the development of an internationally accessible file formats registry, develop a typological framework for classifying data formats and relating them to digital preservation methods.

Approaches to Documentation of Functionality and Behaviour. The preservation of both digital entities and their underlying technologies depends upon representing their functionality and behaviour. Current research had not succeeded in developing approaches to functionality and behaviour abstraction and representation. Efforts to carry this work out depend upon engaging researchers in formalisation methodologies and theory. The DELOS group will focus research efforts to develop methods to describe systems so that they can be reconstructed. It will foster the creation of representations that can be used to establish benchmarks to measure consistency of performance across migrations or emulations. It will promote discussion to develop tests to ensure that it is possible to verify automatically whether or not system behaviour and functionality match that which the application had originally for use in the testbed. The aim will be to produce methods for defining, documenting (including representing) the functionality and behaviour of systems that can contribute to the likely preservation of digital objects.

Enabling the Integration of Digital Preservation Architectures. With some exceptions, a growing body of opinion indicates that the preservation of digital entities can be enhanced if preservation functionality is built into the digital entities or the systems that manage them at the time they are created. For some types of digital libraries this may be feasible, such as those designed with a direct relationship to systems creating the digital objects they will contain, but for most the systems need to cope with the ingest of digital objects in different formats. This means improving our knowledge about what preservation functionality really is and ensuring that this functionality can be effectively communicated to system developers, modelled and implemented by them. Traditional system analysis and design models do not alert developers to the need to ensure that their designs incorporate preservation functionality. A key step to improving preservation strategies would be for 'preservation functionality analysis models' to be integrated with commonly used system analysis and design tools, such as SSADM, to ensure that preservation functionality is incorporated into systems before they are built. This will form a backdrop for creating a relationship between this work and that of those specifying systems and developing them.

From Preservation Models to Implementation. Models such as OAIS are very high level and the process of translating them into workable systems requires substantial effort. Developers of digital libraries need to be able 'to pull off the shelf' designs for the implementation of digital preservation components. While projects, such as NEDLIB and InterPARES, have done much work to extend the OAIS model the tools to ensure its successful implementation are still not complete and easily available. This task will focus on making tools available to system developers to enable them more quickly to develop digital preservation systems.

Designing, Deploying and Managing Digital Repositories. Current research has demonstrated that digital repositories play a central role in the long term curation of digital objects. There are some emerging models as to how these repositories should be developed and deployed. The effectiveness of the repositories have not yet been evaluated. Research areas include the complexity of engineering generic connections to enable newer hardware to communicate with legacy peripheral devices and the definition of automated testing sequences to enable cost effective assessments to be made of the software in the repository to determine whether it continues to function and behave as it was originally designed to do. Repositories also require access to tools to automate curation processes. This Task will lay the foundation for work during the second phase of the project in the area of automating metadata creation, improving selection and ingest functions, and mechanism for handling retention and risk.

1.1.7   Evaluation

Digital libraries need to be evaluated as systems and as services to determine how useful, usable, and economical they are and whether they achieve reasonable cost-benefit ratios. Results of evaluation studies can provide strategic guidance for the design and deployment of future systems, can assist in determining whether digital libraries address the appropriate social, cultural, and economic problems, and whether they are as maintainable as possible. Consistent evaluation methods also will enable comparison between systems and services. Within the proposed NoE, integration activities on evaluation will mainly focus on content and usage of DLs, since there is a significant lack of appropriate methods, tools and testbeds. In contrast, evaluation of management functions will play only a minor role, since this area can draw heavily from corresponding techniques in computer science.

The cluster will work both on evaluation methodologies in general as well as on providing the infrastructure for specific evaluations. More specifically, there are three major objectives for integration activities:

- On the methodological side, there is a lack of established evaluation practices and also little knowledge about existing ones. Thus, the NoE should aim at the role of a clearing house of evaluation methods, tools and testbeds, and support activities for spreading evaluation knowledge.

- Since evaluation is orthogonal to the other clusters of the proposed NoE, the evaluation cluster will keep close contact with all other clusters and provide advice and support for all evaluation issues that would come up in any of the other clusters.

- For evaluation problems that cannot be dealt with by single research groups, the NoE should provide the infrastructure for enabling these evaluations.

In order to reach these goals, the following integration activities are planned:

Survey of Existing Methodology: The NoE will work towards a comprehensive collection of existing evaluation approaches and methods. This collection should be made accessible in different forms, e.g. as a report, as a static website and/or as a dynamic evaluation forum. Possible methods for achieving this goal are the establishment of working groups and workshops of invited specialists in the area of DL evaluation as well as from related fields like e.g. human-computer interaction, information retrieval and information science.

Support to the DL Community: In order to provide evaluation support to the other clusters, the evaluation cluster will support activities for bringing together application specialists and evaluation experts. For this purpose, the NoE will set up the infrastructure for enabling communication between these areas. Besides activities for spreading excellence, the NoE will also set up an electronic forum for enabling communication about evaluation issues. This way, immediate help can be provided for solving simple evaluation problems; more complex problems may lead to the identification of important evaluation issues which require bigger efforts.

Prototype Evaluation Studies: Bilateral prototype evaluations would bring together two partners, one of them an evaluation specialist and the other a developer (from one of the other clusters) needing evaluation for a new type of functionality. Through the exchange of researchers from the two parties, a specific evaluation would be enabled - allowing the evaluator for the application of a new evaluation approach or method, and giving the developer useful feedback about the strengths and weaknesses of his/her system.

DL Evaluation Testbeds and Toolkits: The NoE will create and maintain a collection of testbeds and toolkits for DL evaluation. DELOS under the FP5 has already created a metalibrary of DL testbeds; however, the testbeds collected so far are mainly collections, some of them also comprise a management component, and even fewer involve the usage aspect. Thus, the proposed NoE will target more comprehensive testbeds. For these testbeds, the NoE will aim at achieving

the role of a clearinghouse - where possible - by resolving copyright issues and maintaining the testbeds at a central site.

Organisation of Evaluation Campaigns: Existing evaluation campaigns in areas of crucial importance to Digital Libraries will be supported by the NoE: the INEX initiative on the evaluation of XML retrieval[2] and CLEF, the Cross-language Evaluation Forum[3] Such campaigns not only provide an infrastructure for system testing, producing reusable test-suites for system bench-marking, but are important tools for community building around specific research topics and enable attention to be focused on problem areas and the identification of solutions.

1.2   Programme for Jointly Executed Research Activities

In this section we give an overview of the joint research activities planned by the network in order to realize the integration effort described above in 6.1.

1.2.1   Digital Library Architectures

As mentioned in section 6.1.1, the work of the Digital Library Architectures cluster will address network and basic services architectures that allow integrated access to distributed digital libraries. Thus, the following objectives are addressed:

- Development of surveys that collect the most significant contributions and promises in Digital Libraries Architectures

- Developments of prototype software modules and components for web services, multiple service composition and management, wireless connectivity

- Test of the solutions on a prototype ongoing application.

In order to achieve these goals, the following research activities will be supported by the cluster:

Surveying the State of the Art: Fundamental architectural problems in digital libraries cover the adoption of new networking architectures, definition and adoption of new standards, integration of system components into a cohesive and consistent digital library application and workflow process, integration with new emerging transmission media. The cluster will survey these subjects, identifying emerging solutions, technologies and promising scientific results. In particular the cluster will produce surveys on:

- Replication and Freshness of Data: Mechanisms for digital libraries.

- Security and Certification: Mechanisms for digital libraries

- Network Architectures including Peer to Peer and Grid architectures (differences, common features, synthesis, application scenarios)

- Collection level descriptions to enhance information discovery and assist with information management within service registries.

Development of Peer-to-Peer and Grid Architectures: Peer-to-peer (P2P) architectures allow for loosely coupled integration. Different aspects of peer-to-peer systems (indexes, P2P application platforms, etc.) have to be combined and integrated into an infrastructure for digital libraries. On the other hand, grid computing architectures allow for sophisticated load balancing strategies within a cluster of components. Following the idea of a service grid, and the handling of the control over shared resources, similar concepts have to be integrated into an infrastructure for digital libraries. The cluster will develop a demonstrator system that implements those relevant

---

[2] http://qmir.dcs.qmw.ac.uk/INEX/

[3] http://www.clef-campaign.org

aspects that have been identified as important for the effective exchange of digital library contents.

Development of Service-oriented Architectures and Workflow Management Facilities: The cluster will develop a feasibility study on common protocols for generic service models that provides appropriate descriptions of the available services. Web services are to be integrated as building blocks into digital libraries both to provide access to individual services and to define common services. For applications that utilize digital libraries of autonomous information providers or applications that manage and control the consistency of a digital library, existing services are to be integrated into workflow processes. Both at the application and at the systems level, different aspects of workflow management like self-configuration and flexibility, high availability and scalability, must be included. The cluster will develop a prototype system that demonstrates the feasibility of these solutions in the context of distributed digital libraries.

Development of Mobile Information Components:
The cluster will design and develop a demonstrator that embeds innovative solutions to the most compelling requirements for the access to digital libraries using a combination of wired and wireless connectivity. The demonstrator will embed a specific middleware that will adapt the content to the limitations of handheld devices, for the specific context of digital libraries.

Development of a Medical Information System that exploits Innovative Solutions: The cluster will check all the subjects of investigation referred to in the previous items within the framework of a medical application, currently ongoing at UMIT, and using real data from clinical applications. Researchers from ETH Zürich will guide this evaluation at UMIT. There will also be an exchange of PhD students between ETH, UMIT and OFFIS.

1.2.2   Information Access and Personalization

As mentioned in section 6.1.2, the work of the Information Access and Personalization cluster will address data and user models, algorithms and languages for collection, organization, integration, searching and personalization of information in digital libraries. Thus, the following objectives are addressed:

- Construction of a comprehensive framework for Information Access and Personalization, which can serve as reference point for the DL area.

- Research on new models and methodologies in order to overcome inefficiencies of existing ones.

- Development of toolkits and systems for purposes of re-use and demonstration of proposed methods and models.

In order to achieve these goals, the following research activities will be supported by the IAP cluster (some of which will be in co-operation with other clusters as explained in section B4.1.2 and illustrated in a few indicative cases below).

Development of a Conceptual Model of DL Core Functionality: the services that are provided by a DL depend on the requirements of the community that has set up the DL. Most of the DLs, however, share a "core" functionality, e.g. repository, search, access. So, a model of the core DL functionality will be examined. Such a model must be extensible in order to serve as a reference point for existing digital library systems as well as a starting point for further research activities. (This work will be a joint effort between this cluster and the Evaluation cluster.)

Joint Research on Information Access
Development of Conceptual DL Data and Metadata Models: In order to provide a framework for modeling and manipulation of different data and metadata for accessing data of various types and forms, the IAP cluster will work towards the development of general conceptual models and generic manipulation operations at the data and metadata level for digital libraries.

Development of Query Specification and Processing Schemes: The IAP cluster will work on the improvement of existing models, languages and query processing schemes and development of new ones for information access.

Development of Metadata Models and Algorithms: The IAP cluster will consider more sophisticated representation, query and manipulation schemes for metadata following the general framework provided by the cluster. The purpose of these development activities will be to work towards more efficient and effective representation and exploitation of metadata in digital libraries.

Development of Data and Metadata Generation Algorithms: The IAP cluster will study improvements over existing algorithms for automatic creation and maintenance of data and metadata and will develop new ones wherever possible. A main focus of these development activities will be on generation of richer and more valuable metadata. Data provenance and more general annotations are issues that will be studied in the context of these activities as well. . (This work will be a joint effort between this cluster, the Audio/Visual and Non-Traditional Objects, and the Knowledge Extraction and Semantic Interoperability cluster.)

Joint Research on Information Integration

Development of Conceptual DL Integrated Metadata Models: In order to provide a framework for modeling and manipulation of different metadata for accessing information stored in multiple sources, the IAP cluster will work towards the development of general conceptual models and generic manipulation operations for metadata used for integrated access over federations of digital libraries. This framework may be an extension of the framework for information access in one source in the direction of reacting to problems stemming from information heterogeneity, redundancy.

Development of Integrated Metadata Models and Algorithms: The IAP cluster will consider generation, representation, query and manipulation of metadata focusing on particular problems stemming form the distribution of information. These models may be extensions of models for information access in one source to multiple sources.

Development of Integrated Query Specification and Processing Schemes: The IAP cluster will work on the improvement of existing query specification and processing schemes and development of new ones that will deal with the specific problems regarding integrated access over federations of digital libraries.

Joint Research on Personalization

Development of Conceptual DL User Models: The IAP cluster can start from a collection of existing user models and work towards the construction of more general ones. Identification of a set of general operations that are applicable to certain parts of a model would be extremely useful in the direction of developing a comprehensive framework for Personalization in digital libraries.

Development of user profiling algorithms: Specific models and algorithms for creating and maintaining user profiles that are appropriate and efficient for different DL settings will be investigated and developed under research in the IAP cluster. Expressiveness of user profiles and privacy of users are two conflicting issues that must be taken into account. (This work will be a joint effort between this cluster, the Audio/Visual and Non-Traditional Objects cluster, and the User Interfaces and Visualization cluster.)

Development of Content Personalization Approaches: Development of content personalization approaches will become necessary. The starting point for these approaches will be either the weaknesses of current approaches, or the lack of approaches for new types of applications.

Development of Interaction Personalization Approaches: Development of interaction personalization approaches will also become necessary. Improved algorithms will be developed. The purpose of these activities of the IAP cluster will be towards more intelligent and user-friendly digital library interfaces that will be studied and developed by the User Interfaces and Visualization cluster.

Access to museums and web content by visually impaired people. Research efforts and innovations are to be exploited to improve the capability of blind people to become users of virtual libraries, museum collections, web sites on cultural heritage, and cultural material on CD or DVD. A preliminary demo, already available in the frame of project HELP, is based on a newly developed electronic device to perceive paintings. Further research will be pursued to develop a "web reader" and a multi-sensorial prioperceptive "digital touch device", by integrating the multiple competences of the partners.

The afore-mentioned research activities will be coupled, wherever appropriate, with the following **complementary activities**:

Evaluation of Algorithms: The IAP cluster will cooperate with the Evaluation cluster in order to perform evaluations of proposed algorithms and methodologies. Appropriate evaluation methods and measures will be examined. In lack of appropriate existing ones, possible improvement of existing or development of new ones is planned.

Development of Prototypes: Prototype systems will be implemented in various cases with the purpose of testing, evaluating and proofing the efficiency and effectiveness of new techniques, models etc.

Creation of Toolkits of Algorithms: In order to promote re-use of existing methods, appropriate sets of related algorithms will be assembled into toolkits that can be re-used and integrated in research efforts of the DL community as far as information access and personalization are concerned. The IAP cluster will deal with algorithms that are related to the personalization operations of a digital library. Algorithms that can be developed will be dealing with weaknesses of existing ones and will be based on knowledge accumulated in the NoE.

Development of Applications: Apart for prototypes, systems will be implemented for specific applications that will be made accessible to a broader DL user community. The purpose of these applications is two-fold: i) to promote research efforts of the IAP cluster by using them for the development of real systems and ii) to allow the results of these efforts to reach digital library users.

1.2.3   Audio/Visual and Non-traditional Objects

The integrated research in this area will focus on:
- integrating the state of the art of scientific developments and technologies;

- exploring technology advances;

- achieving a better understanding of the requirements of advanced digital library applications;

- promoting cross-fertilization of the different scientific fields involved.

In particular we will address the following objectives:
- development and integration of advanced scientific results and technologies for information capturing from non-traditional objects;

- development and integration of advanced scientific results and technologies for information management;

- investigation of methodologies for the integration of domains of knowledge and user communities to enhance the functionality of the digital libraries with non-traditional content;

- integration of scientific results and engineering developments for the access of non-traditional content, using different delivery channels.

The following research activities will be supported:

Joint research on Metadata Capturing for Audio.Visual content

Development of Automatic Annotation Algorithms at the Semantic Level. The cluster will study new models and solutions for the integration of automatic information extraction techniques with higher level reasoning algorithms and their combination with manual annotations provided by domain experts. According to this, the cluster will develop solutions for the organization of audiovisual metadata provided by the standards (MPEG7, TV-Anytime, MPEG4, OWL etc) and application domain specific knowledge (according to domain ontologies) in order to automate as far as possible the process of semantic indexing of audiovisual content. The cluster will also focus on advanced "intelligent" user interfaces that can facilitate the integration of the manually provided user's knowledge with information automatically extracted.

Joint research on Universal Access and Interaction with Audio-Visual Libraries

Models of User Interaction: The cluster will integrate models and experimental methodologies from different science fields to study user interaction with the audiovisual information in diverse delivery media using different interaction devices, in order to understand and characterize the meaning of those interactions with respect to the individual user goals as well as to the specific characteristics of the users' community. The objective is to interpret feedback in terms of user preferences so that dynamic profile maintenance is performed automatically by the digital library and used for dissemination. (This work will be a joint effort between this cluster and the User Interfaces and Visualization cluster.)

Multimodal Interfaces and Delivery of Multimedia Content in Diverse Channels: The cluster will develop prototypes of innovative multimodal interfaces and solutions for advanced audiovisual applications to demonstrate the possibility of integration of different aspects, such as access and retrieval modes, delivery modes, interactivity through different devices, user profiles and preferences. To this end the cluster will develop solutions for delivery of multimedia information according to the user's preferences; the cluster will develop solutions granting high quality presentation for the meaningful parts of the multimedia document according to automatically made semantic transcoding, so as to adapt user's requirements about content to the bandwidth and physical limitations of the users' device. The cluster will also demonstrate innovative tools for content based retrieval, browsing, searching and filtering information in different contexts and domains of knowledge, addressing new query and information retrieval paradigms, based on multimedia standards and context and domain specific ontology mechanisms for retrieval, and will analyze cross-cultural and cross-language retrieval.

Models for Multimedia Standard Mappings: The cluster will investigate possibilities for the integration of preexisting or emerging digital TV content specifications (including MPEG7 and TV-Anytime) with broadcasting standards and specifically DVB, in order to develop efficient personalized added-value services on existing broadcasting channels. This will allow digital libraries to act as information providers of additional information to be delivered through the existing broadcasting and point-to-point (ie. Internet) networks and will reference real-time or stored audiovisual content.

Joint research on Management of Audio Visual Content in Digital Libraries

Content Description Models and Standards and Domain Extensions: The objective of the cluster will be to derive new compatible models of domain-specific retrieval language extensions, and to develop innovative prototypes to evaluate these approaches with relevant digital library communities. The cluster will also study software models for structuring and mapping between the multimedia and application-specific metadata which follow different standards and domain specific ontology extensions of the standards. Examples and prototypes demonstrating such mappings will be developed. The cluster will also study new solutions for segmentation and summarization of audiovisual content: In accordance with existing multimedia standards. The cluster will study models and solutions for automatic multimedia content segmentation into both syntactic and semantic units, and summarization at the semantic level, that exploit domain specific knowledge and cater for user interests or suggestions for use. The cluster will develop examples and prototypes to demonstrate this segmentation and summarizations.

Models of Multimedia Databases and Multimedia Peer to Peer computing: The cluster will study efficient data base management of audiovisual metadata, their dynamic indexing and structure adaptation, as well as matching algorithms for audiovisual metadata in accordance with the international standards (MPEG-7, TV Anytime, etc). The cluster will develop innovative prototypes to evaluate the approaches with relevant digital library communities. The cluster will also investigate new models and the impact of developing technologies and emerging scientific results (like grid computing, peer to peer computing and advanced scheduling algorithms for audiovisual content delivery) in digital library applications with large numbers of audiovisual content users.

Multimedia Profiling, Stereotyping, Dissemination: The cluster will develop models for profiling and stereotyping the users of multimedia applications, and will define appropriate dissemination strategies. The cluster will also develop models of user communities and study how these communities can work with the digital library information. It will present models and evaluate prototypes of integration of user and user communities with the digital library content, checking them within the specific application context of education. (This work will be a joint effort between this cluster and the Information Access and Personalization.)

Joint activity on Demonstrators and Testbeds
Development of Advanced Multimedia Demonstrators and Test Datasets: The cluster will develop example domain specific demonstrators in emerging digital libraries for audio/visual and non-traditional objects – such as libraries of 3D static models, libraries of music data, libraries of 3D video, multimedia libraries. – so as to test the acceptability of the new frameworks and techniques by the user communities, as well as to demonstrate the application opportunities of these applications to companies and organizations that work in this field. The cluster will also establish test datasets to facilitate the evaluation of the results of the research.

1.2.4   User Interfaces and Visualization

The challenges which are crucial for the User Interfaces to future digital library systems reach beyond the traditional limits of information system design: embedded and situational usage, casual users, and 3D environments populated with active multimedia objects. Based on tailorable interaction and visualization models, future interface functionality will include options for selection, filtering, linking, ranking, and recommendation of documents, for collaboration with peer groups in various contexts, for re-using and re-interpreting information objects in different contexts. All these modes of operation have to be designed taking into account various aspects such as: user type, user characteristics (such as age, knowledge, language skills, disabled, etc) and their needs. It will be especially important to address the needs of users who are not necessarily end users (such as librarians, maintainers) and users who are from other relevant domains (e.g. archives, schools, museums). The needs of such users have not been well-catered for by existing DL systems. Moreover, DL users with special needs such as the disabled and the elderly will need careful consideration. It will be interesting to consider how information about the user's context can be exploited in making design decisions. Issues pertaining to collaboratories stem from the cases where the DL user interacts with other DL user(s) either as mere DL users or DL user communities/application domain.

The DL system should provide an environment that supports the user regarding all the phases, activities or aspects pertaining to digital libraries, thereby enabling him/her to realize his/her goal in an effective, satisfactory and efficient manner. Such aspects include: searching and browsing, scanning, indexing, information exchange, integration of results/views from various libraries, etc. The process of developing such a DL interface provides not only tremendous opportunities but also resourceful realizations such as methodologies, guidelines, tools, etc.

The proposed network will contribute to these developments by joint activities that will complement ongoing research in European and national projects carried out by the NoE partners contributing to this cluster. As a baseline, we propose to systematically investigate the relationships between users and the contents of a DL system, because the user interface design

must reflect the specific usage requirements. Besides the inputs collected from other NoE clusters, the cluster will source inputs from relevant literature, DL task forces, DL workshops, relevant organizations, and related projects/efforts. Based on the user requirements obtained, this cluster will pursue the building up of a theoretical framework to serve as a developer's resource for designing DL user interfaces.

Relevance Criteria in DL Systems

The consequence of taking the usage situation/context into account results in rethinking the basic assumptions underlying most contemporary approaches to information filtering and retrieval. This will lead to more realistic definitions of "relevance". As empirical evidence shows, the relevance of an information item to the user of an information system is not only governed by topical appropriateness. In addition, the user's relevance criteria refer to the usefulness, novelty, originality, origin, type, and style of the document (this is not an exhaustive enumeration) as well as to its factual content or subject matter. Contextual aspects such as the user's knowledge, tasks, and goals also influence the purpose for which a document is retrieved.

Development of a Relevance Criteria Model. This activity will analyse the results of ongoing European and national projects in order to confirm these initial assumptions. Positive as well as negative experiences with real DL users and DL user communities will be examined. Then a comprehensive model for relevance criteria will be developed and added to a DL User Interface Design Theoretical Framework. The development of the foregoing framework is another activity also carried out by this cluster.

Context Consideration and Exploitation

Taxonomy and specification language for Context Models. Context entails the physical, technical, social, and organizational environments that pertain to DL systems and their usage. It is worth observing that the developments in mobile devices and wireless technologies have made it possible for users to carry out one or many parallel activities, from virtually anywhere at anytime while at the same time interacting with other user(s). Some contexts therefore are such that the users are only capable or willing to pose vague or ill-specified queries and hence the DL user interface should be in a position to provide the required support or assistance. The cluster will carry out studies towards developing a taxonomy of relevant context models. A language specification will then be proposed which shall encompass the pertinent characteristics and requirements of context models that were identified during the development of the taxonomy. The cluster will then model context-dependent DL aspects.

Systematic Analysis of User Requirements

Acquisition of users' requirements. This will start with the acquisition and the study of the requirements of DL users. The cluster will acquire input from relevant literature, workshops, task forces, and other related research efforts/projects. Input will also be gathered from user communities and users from other organizations (such as museums, industry, schools, archives and virtual organizations). Moreover, input from the other clusters of the NoE will be collected. All the input will be consolidated and studied.

Based on the foregoing study, this cluster will then produce a specification of DL user requirements. In order to realise the specification, the cluster will undertake the following activities:

Identify the DL Users and Stakeholders. DL users will be surveyed to understand their needs. It will be crucial to address the needs of users who are not necessarily end users (such as librarians, maintainers) and users who are from other relevant domains (e.g. archives, schools, museums) who have normally been neglected in the existing DL efforts.

Describe user and usage characteristics. User characteristics such as age, education, experience, disabled, elderly, etc. will be categorized and described, as well as the context of DL usage, including users' physical, technical, social, and organizational environments. The issues of mobility and collaboratories will be addressed here.

Identify User Needs, Goals and Requirements. For each of the above characterizations, the goals will be decomposed into corresponding tasks and scenarios. In particular, the recognition of librarians as users will present an opportunity to address their needs which have been largely neglected in the past. Also the requirements of users with special needs, such as the disabled/impaired and the elderly will be studied, to make more real the Grand Vision that "Anyone should be able to receive all information and services they want from any Digital Library, anytime and anywhere, in the most efficient way."

Development of a User Interface Design Framework
A number of ongoing development projects carried out by the DELOS partners can provide input in terms of scenarios, requirements, interface designs, and finally evaluation results, which can and should be reviewed from a more general perspective than the one defined by the specific project providing the input. Design and functions must be linked to DL user requirements. Taking into account the various contextual parameters of the still diverse application domains is the first step towards a theory of user interfaces for DLs.

The first level of the theoretical framework encompasses issues related to the metaphor that is mostly applied today. The contents of the DL are conveyed to the user as a 'collection of information items' which can be searched or browsed. By exploiting effective user interface presentations/visualizations, the user can inspect the items found. While it is obvious that the collection metaphor abstracts from many features that provide orientation and guidance in a real library, it is less clear how to devise a more realistic appearance of a DL. A number of prototypes employ quite different visualizations aiming at suggesting a spatial and/or temporal framework in which users can navigate. However, this is not sufficient to cope with embedded usages, where access to the DL contents is not seen as an isolated activity, but as part of a larger work (education, leisure, business, etc.) context, where interaction with other DL users, editing and annotating DL documents, and proactive services need to be integrated.

The theoretical framework will therefore ultimately also comprise devices for specifying complex DL services such as collaboration support, user agents, immersive or haptic interfaces, and - beyond the simple access modes which rely on the users' initiative such as querying or browsing - proactive system behaviour, e.g. automatic linking, recommendation, explanation, and enrichment. Moreover, the framework should support efficient packaging, exportation and importation of information/library components.

Development of a Design Methodology and Engineering Guidelines. This activity will develop the outlined conceptual framework by means of joint working groups, exchange of researchers, workshops, etc, in order to gather experience and input from as broad a base as possible. Taking into account that future DL solutions will have to provide integrated customisable components that cover the appropriate functionality needed in a given context, the ultimate goal is to develop a design methodology and guidelines that, starting from a generic user interface, will allow to define tailored technical solutions that can be implemented in a given scenario, starting from the users' needs. Thus, we aim at an integrated DL architecture, comprising user and application-oriented functions, e.g. query and navigation features, that can be adapted to the informational needs of users.

DL Lifecycle Support

In order to enable the DL user to be able to carry out any task or activity within the DL lifecycle in a manner that meets his/her goal, various activities will be carried out.

Identification of DL lifecycle phases and association with functional requirements The phases that characterise the development and the usage of a DL system will be identified. The activity will take into account aspects pertaining to the end user and other DL stakeholders such as maintainers, librarians, and archivists. DL aspects such as user goals, system functionality, activities like scanning, indexing, thesaurus support, multimodal support, multilingual support, multilevel user help, searching and browsing, integration of views/results from various libraries and library search tools will thus be analysed and then associated with the identified DL lifecycle

phases. The cluster will also cooperate with the Information Access and Personalization cluster in order to adapt the DL interface to varying user needs.

Analysis of non-functional requirements. Non-functional requirements (such as scalability, platform independence, abstraction) are rarely addressed in the development of most systems. Such requirements are especially crucial for systems with long lifecycles. DL systems fall under the foregoing type of systems.

Multilevel user help. This activity would not only be resourceful to the DL system user but would also help to spread excellence within NoE when carried it out in cooperation with the Virtual D-Lib Competence Centre. This activity will explore the addition of a "Virtual Users' Guide Library" to the services provided by the VDLCC, that takes into account the knowledge of a user characterization and personalization/adaptation requirements (e.g. profiles, preferences, etc). This knowledge could be used to link the user to a particular degree of proficiency with the DL and/or with technology in general. For instance, we can have three levels of proficiency where, in the first one, there are users with a low degree of proficiency with DL or technology in general, (e.g. elderly people); in the second level there are users with a reasonably-good degree of proficiency, and in the third level we can have people who work with DL or are very skilled in the use of information technologies. This activity aims to build a multilevel user support in which every level presents the same information in a different way, using a different language and granularity in the explanation of how the problem can be solved.

Research on Usability Standards and Visualization
Study on interface and visualization principles. This activity will study interface and visualization principles, appropriate interaction styles and universal usability guidelines with an inclination towards the DL context. It should be mentioned that emerging usability issues relating to DL environment (contextual issues e.g. mobility, collaboration) would necessitate the investigation of possibly additional unique guidelines. A study will explore how DL lifecycle phases and aspects (e.g. library contents, catalogue, communities, activities, etc) can be visualized (e.g. as linear structures, as hierarchical structures, as networks, etc) and inline with the gathered principles. Relevant and effective existing visualizations will be exploited and, where necessary, the cluster will develop novel effective visualizations while at the same time taking into account expressiveness and interactivity issues. The issue of how multimedia and text can be used to enhance the visualizations will also be investigated. Usability studies will be carried out to assess the visualizations.

1.2.5   Knowledge Extraction & Semantic Interoperability
The Knowledge Extraction & Semantic Interoperability WP has two key strategic goals:
- To co-ordinate a programme of activities which brings together research excellence from a range of inter-related knowledge engineering and information management areas, and which facilitates the sharing of experience and expertise amongst practitioners from both DL and Grid/computing science backgrounds.

- To explore the potential of new models, algorithms, methodologies and processes in a variety of technical applications, institutional frameworks and cross-sectoral environments, which will lead to the creation of guidelines and recommendations of best practice for dissemination to the widest possible community of interest.

In order to achieve these objectives, this cluster will undertake the following joint research activities.

Open Access to Information and Data Repositories
The development of repositories for materials for the support of research is at a critical stage. In addition to existing information and data archives, the advent of Grid technologies will result in very large volumes of data requiring storage, curation and stewardship. The development of

repositories for learning materials/learning objects may be considered to be at an earlier stage. Some national stores are being created but at the institutional level, many organisations have not considered the implications of requiring to provide access, to manage and preserve these digital assets. Work is needed to produce a mapping of activities and to identify key issues and barriers to success.

A number of scoping studies will be prepared to evaluate the role of information architectures in enabling open access to a range of knowledge repositories in the support of research and learning. These will cover access to primary data, e-prints (subject, institutional, personal), theses, learning objects, multimedia materials and texts.

Whilst there has been much debate on the benefits of open access and self-archiving to the implementation of scientific research, the concept of open access hinges on open metadata exchange facilitated by for example OAI-PMH, to increase the scope and utility of information services. A study is proposed to evaluate the implications of open metadata interchange on social, economic and technical aspects of Digital Libraries and diverse information holdings. These aspects are particularly relevant to the social sciences and arts & humanities, where there is a tradition for individual researchers to work in a more autonomous manner with less collaborative activity with other research groups.

Provenance:
When considering workflows in research and learning & teaching, the authenticity and original source of data and derived data is of paramount importance, in order to provide validity and credibility to the research or learning outcomes. An overview detailing the state-of-the-art in conceptual understanding and implementation of metadata describing provenance in Digital Libraries and which builds on the work of the innovative eBank UK[4] project, is planned. In this context, a technical review of the use of digital identifiers and linking technologies is proposed. (This work will be a joint effort between this cluster and the Information Access and Personalization cluster.)

Semantic Web, Ontologies and Metadata Schema Registries :
There is now an increasing number of developments in the broad area of Semantic Web/Grid technologies, ranging from the development of Semantic Web enabled Web Services to scoping of terminology servers to provide services to distributed digital libraries. There are many aspects to this work area, and this work package will only focus on selected topics based on partner expertise and community requirements.

There is a growing body of work on registries and their use in the publication and validation of metadata schemas. Exploration of various aspects of this topic are proposed and which will include a state-of-the-art overview of the development of metadata schema registries.

The myriad of existing vocabularies both at domain and high level is a major challenge to implementors and users of digital libraries. One particular domain may have multiple vocabularies which are utilised by the communities of interest. We propose a feasibility study to assess the current adoption of high-level vocabularies of subject terms in Digital Libraries. In parallel, we intend to carry out an identification and analysis of a number of domain exemplars in the application of discrete ontologies, and to include an evaluation of their creation, development, management and applicability across sectors. This will be extended with a study to determine the role and feasibility of semantic web services (annotation, reasoning) within Digital Libraries. (This work will be a joint effort between this cluster and the Information Access and Personalization cluster, and the Audio/Visual and Non-Traditional Objects cluster.)

Knowledge Extraction:
Finally, the creation of data and information repositories of increasing size and complexity e.g. in astronomy, bio-informatics, requires new tools and approaches to analyse, model and render the data in order to derive further information and knowledge for research purposes. This topic will be

---

[4] EBank Uk Project http://www.ukoln.ac.uk/projects/ebank-uk/

explored initially with a study to determine the requirements for and usage of extracted knowledge for bibliometrics, domain analysis, issue tracking and community modeling. It is planned to extend this application area more widely to consider other domains and to implement a comprehensive study which describes the application of text-mining, data mining and other algorithms within diverse digital libraries.

1.2.6   Preservation

Integrated research in the preservation cluster will provide the methodological framework and theory for ensuring that digital libraries research addresses preservation issues and digital libraries incorporate preservation elements in their designs. The research agenda in digital preservation is very broad and the objective will be to tackle a small number of aspects at this stage:

- The establishing of file format registries and the relationship between the typology of file formats and preservation methods,

- The definition of preservation functions as they need to be incorporated into the digital library,

- The establishing of testbeds and validation metrics to test preservation strategies and the preservation worthiness of digital library implementations,

- The enhancement of methods for evaluating materials for ingest into a digital library and for ingesting materials, and

- The definition of guidelines for establishing, auditing, and certifying digital repository functions of digital libraries.

These objectives can be achieved through establishing:

Digital Preservation Testbed Forum.   The Forum will lay the foundations for a digital preservation testbed environment by establishing the framework for testing and validating digital preservation methods, repository implementations and creating metrics for comparability between testbed environments as it is unlikely that a single testbed will meet the needs of the entire community. It is however essential that the outcomes from different testbed strategies be comparable.

Designing, Deploying and Managing Digital Repositories. The digital library community needs access to tools to assess how effectively and efficiently digital library implementations provide preservation functionality. The focus of this activity will be on defining digital repository frameworks and mechanisms for validating the suitability of digital repository implementations. Without such tools it is impossible to audit and certify repositories and therefore to assess whether or not Digital Library implementations achieve any measure of sustainability.

File Format Data and Typology will contribute to the development of file format registries and the mechanisms for their use (e.g. not unlike authority files) . The key activity will be to define the relationship between file format types and preservation methods and to assess the viability of producing generic metrics to measure the viability of this preservation approach. This will be done by creating a file format classification framework with rules for extending the framework to enable it to support the classification of other and emerging file formats. The Preservation Cluster must ensure that the European Community, through DELOS, owns a share of file format registry development, management and access processes.

Documentation of Functionality and Behaviour Metrics. While it is evident that formal methodologies for abstracting functionality and behaviour of digital entities are essential, current research has not defined them. In the first of a two stage process, the Preservation Cluster will establish a workgroup to define what kinds of functionality and behaviour metrics are required if digital libraries are to verify automatically whether or not system behaviour and functionality

match that which the application had originally, previous to its migration, emulation, or retargeting.

During this phase we will begin work on the following:

Enabling the Integration of Digital Preservation Architectures. With some exceptions, a growing body of opinion indicates that the preservation of digital entities can be enhanced if preservation functionality is built into the digital entities or the systems that manage them at the time they are created. For some types of digital libraries this may be feasible, such as those designed with a direct relationship to systems creating the digital objects they will contain, but for most the systems need to cope with the ingest of digital objects in different formats. This means improving our knowledge about what preservation functionality really is and ensuring that this functionality can be effectively communicated to system developers, modelled and implemented by them. Traditional system analysis and design models do not alert developers to the need to ensure that their designs incorporate preservation functionality. A key step to improving preservation strategies would be for 'preservation functionality analysis models' to be integrated with commonly used system analysis and design tools, such as SSADM, to ensure that preservation functionality is incorporated into systems before they are built. This will form a backdrop for creating a relationship between this work and that of those specifying systems and developing them.

From Preservation Models to Implementation. Models such as OAIS are very high level and the process of translating them into workable systems requires substantial effort. Developers of digital libraries need to be able 'to pull off the shelf' designs for the implementation of digital preservation components. While projects, such as NEDLIB and InterPARES, have done much work to extend the OAIS model the tools to ensure its successful implementation are still not complete and easily available.

Assessment of Current Research: In support of an extension of the activities led by ERPANET the Digital Preservation cluster will foster the evaluation of research into digital preservation to ensure that high quality outcomes are widely visible and adopted.

1.2.7   Evaluation

Integrated research in the evaluation cluster, will address both evaluation methodology as well as application-type specific evaluation methods and testbeds. Thus, the following objectives are addressed:

- development of a comprehensive theoretical framework for DL evaluation, which can serve as reference point for evaluation studies in the DL area..

- research on new methodologies will be supported in order to overcome the lack of appropriate evaluation approaches and methods,

- development of corresponding toolkits and testbeds for DL evaluation in order to ease the application of evaluation methods, .

In order to achieve these goals, the following research activities will be supported:

Development of Conceptual DL Models: In order to deal with the complexity of DL systems, appropriate conceptual DL models will have to be developed. Although some preliminary models have been described in the literature, there is still a lack of a comprehensive model describing the structure and behavior of DL system. In order to consider the large variety of viewpoints on DLs among the stake holders in this area, the development a conceptual DL model will be a collaborative effort. Major building blocks of such a model will be classification schemes for content and usage. For the latter, there are a number of different schemes. For example, the 'digital

life cycle' [Paepcke 96][5] focuses on different usage phases: discover sources, retrieve objects, collate objects with respect to some scheme, interpret the content of objects, and combine the outcome with own findings and re-present the result in the form of a new DL object. Orthogonal to this scheme, one can distinguish different levels of granularity, from a simple `move' over a `tactic' and a `stratagem' to a `strategy' for satisfying a complete information need [Bates 90][6] In the context of evaluation, such a model forms the basis for relating the evaluation of different aspects of DLs to one another. At the present time, evaluation of different aspects of DLs proceeds largely without reference to the results of evaluations of other aspects. This is often the case just because DLs are so complex, and have so many different parts, each with their own special investigative techniques, evaluation measures, and so on. How to relate the results of evaluation of the different parts to one another is in itself a significant research question, which needs to be addressed if we are to be able to evaluate DLs as whole systems. This activity will be integrated with an activity of the IAP cluster aimed at producing a conceptual DL model on the basis of a study of user requirements.

Development of New Evaluation Approaches: Starting from the weaknesses of current approaches and the lack of approaches for new types of applications, new evaluation approaches will be developed. A general problem of current evaluation approaches is the insufficient participation of users in the evaluation cycle. Going beyond such techniques as user-centered design, or participatory design, (potential) users of DLs will be enabled to be significant evaluators throughout the design/evaluation cycle.

Development of DL Evaluation Methods: Based on the framework developed, standard evaluation methods for DLs can be developed. This work can start with meta-analysis of existing studies (including what has been the effect of evaluation), in concert with, or followed by, collaborative efforts to compare and evaluate techniques for evaluating DLs (including multiple evaluations of the same system). Based on the results of these activities standard techniques, methods, measures for DL evaluation can be developed and be made available to the entire DL community. For new types of content or usage, the application of existing evaluation approaches also will require the development of appropriate evaluation methods. For this purpose, joint research of evaluation experts and application specialists will be required.

Development of Evaluation Toolkits: Given a set of standard evaluation methods, there may not be appropriate toolkits for applying these. Especially, tools for collecting and analyzing experimental data should be developed in order to ease the application of the corresponding methods.

Creation of Test-beds for New Content and Usage types: With regard to new types of usages, the current CLEF and INEX test-beds will be used as starting points for including new usage types for the underlying content, such as e.g. searches involving controlled vocabulary, question answering or Web search on multilingual content in the case of CLEF or interactive retrieval and browsing of XML documents within INEX. The definition of these new usage types will be based on appropriate user studies. For new types of content, efforts will be undertaken for creating appropriate test-beds once there is sufficient interest from the research community. Possible candidates in this area are MPEG-7 or TV-Anytime audiovisual documents, 3D data or application-specific data like e.g. from the medical or the geo-satellite domain.

Test-Beds for Usage-Oriented Evaluation: The evaluation of interactive or collaborative use suffers from the lack of appropriate test suites. Thus, different approaches for developing test-beds for usage-oriented evaluation will be pursued:

- Starting from existing test-beds (e.g. in CLEF or INEX), methods for exploiting these test suites for usage-oriented evaluation will be investigated,

---

[5] Paepcke, A.: Digital Libraries: Searching Is Not Enough - What We Learned On-Site. D-Lib Magazine 2(5), May 1996.

[6] Bates, M. J.: Where Should the Person Stop and the Information Search Interface Start?.Information Proc. & Mgmt. 26(5), pp. 575-591, 1990

- Another possible approach is the creation of a test-bed of user interactions in DLs. Such a test-bed could be constructed in a distributed manner, by different groups reporting the results of separate studies to a central agency, which would maintain the test-bed. Alternatively, publicly available demonstrators could be developed, from which appropriate data could be collected. The data collected in either of the two ways would allow investigation of a variety of issues by many different groups, without their having to record data de novo.

- As a third approach, a test-bed framework will be created, comprising both content and management components (a modular DL system). Using this framework, research groups could plug in their own services. For evaluation, either the database of user interactions could be used, or the resulting system would be evaluated with new users. Examples of such frameworks are Daffodil[7] or systems based on the SDLIP[8] interoperability protocol.

1.3   Spreading of Excellence activities

In order to coordinate the dissemination effort of the wide range of activities carried out by the Network, a Virtual D-Lib Competence Centre (VDLCC) will be established. In addition to providing support to the dissemination of the Network activities, the centre will provide education, training and technology transfer to research, memory institutions and industrial organizations in the field of Digital Libraries. The Virtual D-Lib Centre will be implemented by the coordinated efforts of three institutions participating in the Network, strategically located in Europe: CNR-ISTI in Italy, UKOLN (University of Bath) in the UK, and Netlab (University of Lund) in Sweden.

Some activities of the Virtual D-Lib Centre will be decided autonomously by the VDLCC management, on the basis of the needs and requirements of communities external to the Network, while many other activities will be executed in support of the needs and proposals from the Network clusters. It is expected that most of the scientific dissemination activities of a cluster (typically workshops and summer schools) will be coordinated by the VDLCC, which will provide administrative, secretarial and logistic support and will ensure that there will be an inter-cluster coordination of those events. The activities of the Virtual D-Lib Centre fall into different categories, based on the main objective of the activity, the contents and the targeted recipients.

1.3.1   Scientific Dissemination

European Conference on Digital Libraries (ECDL). The ECDL series will be supported. The conference will be organized each year in a different European country and representatives of the most active research teams in Europe, plus key figures in the Digital Library community from the rest of the world, will be involved in the preparation of its scientific programme. From a technical point of view, the ECDL goals are to define those digital library parameters which especially influence issues of access to, retrieval of, and interaction with information; to identify key problems which must be solved to make digital library service an effective reality; to identify a general structure or framework for integrating research and solutions; and to propose and encourage specific, high-priority research solutions within such a framework.

A Series of Thematic Workshops. This series of workshops will give the opportunity to European researchers to present results of on going research activities and to exchange opinions and experiences in an informal friendly environment. Position papers, progress reports on projects in progress and demos of running prototypes will constitute the main form of communication in these workshops. The workshops could also provide course material for the Summer School in

---

[7] http://www.daffodil.de

[8] http://www-diglib.stanford.edu/~testbed/doc2/SDLIP/

Digital Libraries and a platform for communicating and disseminating standardization issues. Each workshop will have a Programme Committee responsible for the scientific quality of the workshop communications. The cluster coordinators will propose the specific themes of each workshop and will remain responsible for the programme and its scientific quality, while the VDLCC will provide all the needed organizational support, such as preparation and dissemination of the Call for Papers and Call for Participation, pre-registration of participants, venue, publication and dissemination of the proceedings, etc. Moreover, the VDLCC will ensure a close coordination between the clusters, to avoid overlapping themes, conflict of dates, too many events in a short period, etc. The exact number of workshops to be organized each year will be based on the detailed work plans defined by the Work Package leaders.

A Series of Brainstorming Workshops. A series of brainstorming workshops will be organized. The objective is to define "future research direction" reports in the field of Digital Libraries and to provide valuable input for the definition of future research programmes in the domain of Digital Libraries both for the EC and national research funding agencies. Top researchers from all the parts of the world will be invited to actively participate in these workshops. This series of workshops will have no pre-defined periodicity, as it is expected that each workshop will be triggered by a specific need raised by one or more of the clusters.

1.3.2   Education and Training

A Series of Summer Schools. A series of Summer Schools will be organized. The Schools will provide high-level courses on the domain of Digital Libraries and its underlying technologies. Moreover, they will offer a unique opportunity for a complete update on all aspects of Digital Libraries and will encourage contacts between participants, lecturers and interested parties. The schools are directed to members of the research community (in the wide sense): primarily graduate students, but also young researchers and professionals involved in R&D in Digital Library related areas. A Scientific Committee will oversee the organization of the Schools to ensure the effectiveness of the curricula and the quality of the Summer School, which will typically be of one week duration. The criteria adopted for the organization of the Summer Schools will mirror those of the thematic workshops: the cluster coordinators will be responsible for proposing the Summer School topics and for the detailed programs, and the VDLCC will be responsible for providing all the necessary organizational support.

An International Inter-university PhD degree on Digital Libraries. The Competence Centre will start proactive actions with some of the academic institutions participating in DELOS, in order to establish agreement between them for the purpose of defining and granting a PhD in Digital Libraries. That should provide graduate students from different disciplines a common view and understanding of the main Digital Library technologies and applications. Students will be appointed a main supervisor from one of the institutions that have agreed to the program, and the students will have the possibility of spending periods of time in different institutions. The final PhD title (or equivalent) will be granted by the institution of the main supervisor.

Research Exchange Programme. In cooperation with the cluster coordinators, a research exchange programme will be implemented. This will take the form of the exchange of scientific visits and stages between members of the NoE research teams. However, as the domain of Digital Libraries is application-oriented, the exchange of visits between researchers from academia and industry and other application environments (cultural heritage institutions, broadcasting, electronic publishing, etc.) will also be organized. Whenever possible, the Network will try to take advantage of the Marie Curie programmes for the mobility of human resources.

Preservation Support to the Digital Library Community: The Digital Library Community will benefit from improved understanding of the place that preservation plays in the development of effective, efficient and sustainable digital libraries. Outreach, data dissemination, and communication activities are planned to make this happen. Key among these will be an "Online Digital Preservation Educational Service", to establish the framework for a "Digital Preservation Educators Forum", which can deliver during the later phase of the project educational material online, to train students, researchers and professionals in current digital preservation strategies and

thinking. The eventual aim of this activity is to define, develop and deploy online learning services in digital preservation for digital libraries in a multilingual environment. The results of research activities in preservation need to be made accessible, in order to reduce the redundancy of activity that is currently characterising the digital preservation community.

1.3.3   Technology Transfer

A Series of Awareness Events. In cooperation with the cluster coordinators, a training programme will be run consisting of a series of awareness events (courses, tutorials, demonstrations, etc) to be organized in collaboration with specific application and industrial communities. These events will present Digital Library technologies and successful case studies in a customized manner to user communities (Electronic publishing, Libraries, Archives, Museums, Broadcasting, etc.) and to interested potential industry partners. In addition, they will present new prototype systems to particular communities and demonstrate how they could improve the services provided or their productivity - depending on the type of community involved. The events must be focused on specific application/commercial environments, on their demonstrative mission, and on their window of opportunity. 2-3 training events will be planned annually. Each training event will be replicated at the three sites of the Virtual D-Lib Centre (University of Bath-UK, CNR-ISTI-Italy, and University of Lund-Sweden) in order to better cover the European space. The instructors of these training courses will be sought in the NoE Digital Library community as well as in the worldwide research community. Considerable contribution to the preparation of these courses is expected to be provided by the NoE research activities.

A Series of National/Regional Technology Transfer Workshops. The members of the NoE will establish contacts with their national application communities and industry, for the purpose of organizing joint workshops on themes of mutual interest. Particular attention will be given to the involvement of SMEs in each country. The NoE will sponsor a series of such workshops by funding the participation of a number of invited speakers for each workshop.

1.3.4   International cooperation

A Series of Bilateral Workshops between European and non-European countries. The objective is to establish collaborative actions with other research communities outside of the European Union, in order to strengthen those collaborations already established by the FP5 DELOS NoE (with the US and the Russian Digital Library communities) and to extend them to include other countries (for example, the countries of the Mediterranean area, Canada, Australia, Latin America), with which the European Union has in place an agreement for scientific cooperation under the 6th FP.

Use of Marie Curie Programmes for International Fellowships. The Virtual D-Lib Centre will capitalize on the Marie Curie programme, in which many actions are open to third countries, to attract researchers from non-European countries to participate in the research activities of the Network and establish in this way possible cooperation links with countries outside of Europe.

1.3.5   Dissemination and Visibility

Electronic Newsletters. The Network will publish a regular (two-monthly) electronic newsletter. In addition to information and reports on DELOS events and on related activities of importance to the digital library community intended in the widest possible sense (i.e. the research, user and application sectors), the newsletter will provide a forum that can be used by key players in the area to express their views on DL-related topics of interest. Induividual clusters may decide to publish more specialized e-Journals, and in this case the Virtual D-Lib Centre will act as the distribution channel.

The Network will also prepare and maintain a Network fact sheet on CORDIS and the Digicult site. This will be checked at regular intervals (min. 6 months).

In addition, the Network will deliver to the Commission an A4 "milestones" page every two months for inclusion on the Commission's Web site and in the eCulture newsletter. It will contain

specific and digital library specific issues and progress and will be written for the non-expert audience.

Network Web site. A Network Web site will be designed and implemented. It will maintain and make accessible the collection of all the deliverables of the Network, management reports, demos etc. In addition, the Web site will provide information on the latest research results in the field of digital libraries as well as the latest information about international projects, initiatives, conferences, etc in the digital library domain. It will also provide access to all major Web sites related to digital libraries. In essence, it will constitute a portal to the international scene of digital libraries and an important part of the Network brand. The Web site will also contain a "citizen charter" which will provide information about the Network rational, objectives, expected results, partners, funding, etc. The citizen charter will be written in such way to be easy to read and to understand. The number of visitors as well as the reference of the Web site by all the main search engines will be considered as important success indicators.

1.4 Management Activities

The management structure of the Network (described in detail in Section 8) consists of a set of independent organs, interlinked to ensure an overall cohesion of activities.

The Advisory Board. The role of the Advisory Board is threefold; it will provide: i) external assessment of the results achieved by the network, acting as a peer-review council, ii) guidance with respect to the network's strategic orientations, iii) support of the network's excellence. The Board will be composed of 5 senior external experts in the field, able to provide an impartial assessment of the work carried out by the network.

The Scientific Board. The Scientific Board receives guidance from the Advisory Board and will provide a high level supervision of the network's integration and research activities. It will promote and assess the scientific quality of the network, will define means to implement all major technical decisions, will define funding policies, will identify the need for significant changes in the NoE work-plan and will oversee their implementation. The Board is composed of the Scientific Coordinator of the Network and the Work Package Leaders. The Administrative Coordinator will also attend meetings of the Scientific Board to guarantee coordination between scientific and integration activities and the necessary administrative back-up actions.

Work Package Steering Committees. For each work package (i.e. the detailed activities planned for a cluster over an 18 months period), there is a Work Package Steering Committee, which receives guidance from the Scientific Board and is responsible for overseeing the integration and specific joint research actions to be undertaken by that work package. Each Steering Committee is composed by the main contact person of each institution participating in the Work Package and is chaired by the Work Package Leader. Each Work Package Leader nominates one or more members to act as liaisons with the other Work Packages, in order to promote and coordinate integration between work packages. Each Work Package in general has one or more Tasks of joint research, integration and dissemination activities. For each task, a single research group is usually responsible for both the research and the corresponding integration activities, in consultation with the leaders of the groups participating in the task. Different types of research, integration and dissemination activities may take place for each area: studies, workshops, evaluation of algorithms, prototypes, toolkits, applications, summer schools, etc. Whenever the nature of the activity requires it, the responsible group, in agreement with the Work Package Leader, will appoint an individual to take the main responsibility of organizing and directing the specified activity, in cooperation with the Virtual D-Lib Centre, if appropriate.

Virtual D-Lib Centre Management Committee. The Virtual D-Lib Centre Management Committee receives guidance form the Scientific Board, and is responsible for supporting all the activities to spread the excellence of the Network, i.e. scientific dissemination, education and training, technology transfer, international cooperation, dissemination and visibility.   The

Committee is composed by a representative of the institutions implementing the Virtual D-Lib Centre (currently CNR-ISTI, UKOLN, NetLab), and is chaired by the Scientific Coordinator of the Network. In close cooperation with the Work Package Leaders, the Committee will plan and organize the specific events in support of the Network

Specific Task Forces. The Scientific Board can establish Specific Task Forces to address particular transversal activities to satisfy identified needs of the Network and/or to provide additional support to the activities of the work packages. A Task Force will be composed of a Task Force Leader and of dedicated members appointed by the Scientific Board.

The Members General Assembly. The general assembly of the Network participants will provide a forum to be used by all the Network members to initiate discussions and express opinions regarding both research and integration activities. It will also provide the Scientific Board with a broad audience to present the main directions and intentions of the NoE and from which to receive input. The General Assembly is a virtual assembly of all the Network participants, existing through a moderated discussion list. The Scientific Board will appoint the moderator of the discussion list.

1.5   Plan for using and disseminating knowledge

In this section are described the plans for the management of knowledge and intellectual property, and the plans to ensure the broadest possible usage of the Network results.

1.5.1   Intellectual Property Rights

Use or dissemination of the knowledge must be detailed to ensure the exploitation of the results in a fair and durable way. The principles below are detailed in a more complete manner in the Consortium Agreement and in the Annex II, General Conditions of Contract.

Know-how (pre-existing knowledge)

- To be integrated as an appendix to the CA, each participant will identify and will transmit to the administrative and financial co-ordinator, the list of :
    - o the pre-existing know-how it is providing to the project
    - o the pre-existing know-how it explicitly wants to exclude from the other participants' access.
- It is agreed that access rights to the pre-existing know-how shall be only granted on written request.

Knowledge (produced by NOE)

- The knowledge produced during the project shall be the property of the participant carrying out the work leading to that knowledge. If several parties have carried out the work leading to the production of such knowledge, they shall respectively have joint ownership on the knowledge produced.
- The parties are expected to protect their knowledge as efficiently as possible in accordance with Annex II – General Conditions II.20.1

Access rights

- Access rights to pre-existing know-how needed for the execution of the NOE work are granted on a royalty free basis.
- Access rights to pre-existing know-how for use of own knowledge should be granted on fair and non-discriminatory market conditions for commercial use and on royalty free conditions for non commercial research.

- Access-rights to knowledge for Use of own knowledge for commercial activities should be granted on favourable conditions. For educational and research use they should be granted on royalty-free conditions. In any case, a separate agreement aimed at specifying the terms of use should be signed.

- Parties that may be subject to third party obligations must inform the members of the consortium.

- The participants are strongly encouraged but cannot be forced, to grant to other participants

  - on a royalty-free basis access rights to their own Pre-Existing Know-how for the execution of the project

  - at favourable conditions access rights to their own Knowledge needed for the use of knowledge produced by other participants.

Publications

There is a clear balance to meet between the interest of publishing and the intellectual property protection. The Consortium Agreement and the Annex II, General Conditions of Contract, implement a mechanism aiming at providing all parties with information related to the forthcoming publications. If the publication infringes one some confidential information, the party wishing to make the publication may delay or modify the content of his/her publication.

1.5.2 Industry involvement

Industrial developers have not delivered DLMS of sufficient richness, robustness, and sustainability to meet the needs of the potential digital library communities. Recent reports of evaluations of digital library products by National, academic, public, and commercial organizations all reach the same conclusion: "none of the products on the market provide the services necessary to implement a sustainable digital library service of any diversity or scale." Those requiring digital library implementation mechanisms can identify the shortcomings of the products, but do not know what an adequate product would look-like and what services it would support. The commercial developers are in the same position. A strength of the DELOS Network is to have pulled together researchers who have begun to demonstrate what digital library services should look-like and how they can be put together. The Network will take this forward in four ways:

First, very close relationships will be established with those Integrated Projects funded by FP6, which are related to digital libraries. We are aware of at least two proposed IPs currently under consideration for FP6 funding, with relevant research focus and suitable industrial partners. Projects funded under FP5, related to Digital Libraries, will also be considered, as several of these projects have participating research organizations that are also participants in the DELOS Network. Through these relationships DELOS will contact and involve industrial partners who are interested in digital library technologies and systems in order for them to become participants in the DELOS FP6 Network.

Second, a series of "Digital Libraries All Project meetings" will be organized which will involve many digital library projects funded both by the EU and by national initiatives. This type of meeting facilitates the building and solidifying of relationships with a large number of industries involved in digital library R&D activities and will allow us to identify not only potential partners, but also to receive input from them. We experimented with this type of meeting during the FP5 DELOS NoE with considerable success.

Third, we will target a series of 'Awareness' workshops and tutorials for industrial developers, addressing in particular SMEs. These workshops will disseminate the research results of DELOS and other EU-funded projects, offering the industrial participants the opportunity to evaluate the benefits of Digital Library technologies. We are planning to have a two-step process to help identify the most interested industry partners. First, a series of Awareness Workshops will be

organized at the national/regional level, with the assistance of the local Network participants for the selection of the most promising SMEs and other industries to be invited at the workshops. We will then organize an International Awareness Workshop, where the top 4 or 5 SMEs and other industries that have expressed interest in each national workshop will be invited, and where we will start discussions for a more concrete involvement in the Network activities.

Fourth, we will increase the interest of industrial partners by raising the awareness of digital library technologies among national, public and private organizations, interested in delivering usable digital libraries services, and seeking technologies able to provide functionalities and behaviors at the level required by the services they intend to offer. We will do this in part through seminars, encouraging these organizations to participate in our user needs studies, and in part through our web-based services and newsletter.

1.5.3   User involvement

The involvement of the application communities (the end users) will be of course crucial, as their feedback will be needed to ensure that the research activities of the Network will address their needs. In addition, the involvement of professional roles in Digital Libraries will also be crucial, to ensure that the capabilities of the management tools being developed will address their needs. Therefore attention will be paid not only to the role of the end users, but also to the role of other DL professionals such as librarians, archivists, content providers and maintainers.

The Network intends to build on experience gained by the FP5 DELOS NoE, where a joint DELOS/NSF Working Group on "Reference Models for Digital Libraries: Actors and Roles" was set up with the purpose of better identifying actors and their roles within the context of digital library applications. Actors in a digital library were defined as persons, organizations, automata or any kind of entity in general that play roles in the production, dissemination, management and use of digital information and services related with the digital library.

From this abstract point of view, a digital library can be defined simply as a set of services, managed by one or more classes of professionals (special classes of users), conceived for the benefit of one or more classes of users (end users), and possibly related to external classes of agents (special classes of users).

A User can be any class of actors that might interact with the digital library's services (scholars, citizens, employees, etc.). A DL Professional can be any generic actor responsible directly or indirectly for the set up, maintenance, and evolution of the digital library (cataloguers, indexers, or reference professionals, managers or decision makers, new emerging professionals such as computing staff, etc.). An Agent is any class of external actor that plays a role in the conception, implementation, maintenance, optimization, improvement or augmentation of the digital library's services (publishers, authors, contributors, etc.). We foresee four main mechanisms for the recruitment of these different classes of users in the Network activities.

First, the D-Lib Center will allow the Network to contact a large number of users belonging to all the above classes through its extensive technology transfer activity. Second, we plan to appoint representatives of many cultural institutions (libraries, archives, museums, broadcasting) as new partners of the Network. These representatives will act as liaison between the Network research community and the communities that they represent. Therefore, they will be able to attract into the Network activities significant and diverse user classes. Third, the ERCIM community will be involved. ERCIM is a large research consortium, which includes the major research organizations in the IT sector (presently 16 major national and international organization). This community will provide representatives of scholars and authors. Fourth, larger communities of users will be reached through exploitation of the portal, the other web sites and the various events organized by the Network.

The type of user involvement will depend on the type of activity. In the user requirements acquisition phase, appropriate requirements will be derived through an extensive use of questionnaires and interviews. In the experimentation phase, a direct involvement of the users will be experimented. In particular, users will be directly involved in system testing. This will take

place mainly in the context of system demonstrations that will be organized in collaboration with museums and reference institutions. Examples of possible venues for system demonstration are the "Palazzo Medici Riccardi" and the "Sala d'Armi di Palazzo Vecchio", two historical buildings in Firenze with which one of the Network participants already has in place a cooperation agreement. Users will also participate in the evaluation of interaction meaning in various types of interfaces. For example interactions with a TV controlling device will require cognitive models for the users to be built in order to infer what was the meaning of the interaction (if the user liked or disliked the movie, etc). Finally, in the dissemination phase, appropriate user profiling will be extracted again by extensive use of questionnaires and interviews.

1.5.4   User profiling and users with special needs

Universal access to information is clearly part of the Grand Vision ("…any citizen to access all human knowledge…"), and that of course includes access to information, resources and services by special types of users. The User Interface cluster will characterize the various types of users from several points of view, including age, educational level, cultural background, users' abilities and disabilities. This characterization will be part of a "user profile", and the combination of different types of profiles can provide the basis for an exact definition of the user needs and assist in determining which services and resources might be delivered to individual users and how they would be presented. User profiling can take different meanings depending on the type of users involved. For Digital Library professionals (administrators, librarians, archivists, etc) the meaning is mostly related to the role that those users have in the interactions with the Digital Library, while for end users the meaning is mostly related to information filtering and presentation.

From the technical viewpoint, some of the planned research activities will lead to the development of tools to transform retrieved documents into formats more accessible to users with special needs (e.g. multimodal interfaces). The commitment of the Network to inclusive approaches to the specialized needs of DL users is reflected for instance in the joint research activities planned in the cluster on Information Access and Personalization which aims to provide access to content for the visually impaired.

As the Network is widened, it will be broadened to include researchers, experts and other actors to ensure that the development of DL methods, technologies and services meet the needs of the broad range of user communities. DELOS is committed to ensure that development efforts by current members and its extension to include new players will help to lower existing barriers to access digital library services by all possible users.

1.5.5   Involvement of cultural institutions

Digital Libraries cover a wide spectrum of cultural institutions and cultural areas. The Network will be expanded to cover the following types of cultural institutions:

- Libraries
- Archives
- Museums
- Broadcasting Agencies

One or two very representative institutions in each category, in selected cultural areas, will be invited to become partner of the Network. Their role will be to act as a liaison between their community and the digital library research community, assisting in particular in the definition of the dissemination activities of the Network. This will permit the identification of a customized dissemination strategy, allowing well-defined technology transfer actions to be carried out for specific cultural communities. This approach has already been used during the FP5 DELOS NoE, leading to the establishment of very good relationships with IFTA (International Federation of Television Archives), ICA (International Council of Archives) and IFLA (International Federation of Library Association). The cooperation with those international organizations will help us to

establish contacts and identify the most appropriate partners through mechanisms very similar to those described for the involvement of industrial partners.

The areas that will be considered with high priority when expanding the Network are the following ones, due to their cultural, scientific and social value; in addition, due to their requirements and characteristics they would allow testing most of the technologies developed by the Network.

- Arts and Humanities, for the wealth of artistic and historical contents, analysis of content quality, definition of strategies and techniques for content communication, definition of models for representation of user characteristics.

- Medical, for the definition of strategies and techniques for distributed data, for reliability and quality of service, for various classification, clustering, abstraction, and summarization techniques, for context (location, time, situation) based relevance and relevance feedback, for data mining algorithms, for security and privacy of access.

- Juridical, for the assessment of legal rights related to the duplication, dissemination, and disclosure of cultural heritage materials as well as the "analysis" of issues related to the protection of privacy in information/communication systems.

1.5.6   Cooperation with other EU-funded initiatives

Given the broad spectrum of DELOS activities, it will be mutually beneficial to establish cooperation links with other initiatives funded by the European Union. In the following section we describe the cooperation links that are presently established with other initiatives.

1.5.6.1   BRICKS Integrated Project

Many of the DELOS activities have some relevance also for the activities planned in the BRICKS Integrated Project, selected for funding under FP6. Given that 5 pparticipants in DELOS are also participants in BRICKS (namely CNR-ISTI, UOA, UNIFI-MICC, FHG/IPSI, CF), it appears natural to consider those organizations as the bridge between the activities of the two initiatives. A fruitful cooperation can be established in all the 3 areas of work in which BRICKS project is structured, namely the infrastructure, the service, and the sustainability area.

- At the *infrastructure* level, DELOS will provide results and insights from the research community, while obtaining from BRICKS feedback on the applicability and effectiveness of these research results, as well as requirements for new basic technology.

- At the *service* level, BRICKS will collect and bring to DELOS the user experience on the specific services implemented by the project, in order to stimulate a debate on more advanced methods for tackling the involved problems; the DELOS community will bring new ideas coming from the research world to test them against experts of applications in the Cultural Heritage area.

- At the *sustainability* level, BRICKS will exploit the technology transfer activity of DELOS in order to disseminate its technology and services to identified regional communities during Awareness Events, with special emphasis on SMEs.

The tables below indicate the relevance of the DELOS activities (where the partners in common are involved, and are identified in the first row by their participant number) to the activities (bricks) of BRICKS. For convenience, we have included also a table with a short description of the components of BRICKS.

Relevance between DELOS Integration Activities and BRICKS components

DELOS Cluster	Integration Activity	2	5	7	8	40	BRICKS components
ARCH	*Peer-to-Peer Architectures*	♦	♦			♦	All core bricks

DELOS Cluster	Integration Activity	2	5	7	8	40	BRICKS components
ARCH	*Service-oriented Architectures*	♦	♦			♦	All core bricks
IAP	*Data and Query Models; Query Processing Schemes*	♦	♦				I, Coll, SB
IAP	*Metadata Models; Metadata Query Processing Schemes*	♦	♦				MM, I, Coll, SB
IAP	*Metadata Manipulation*	♦	♦				MM
IAP	*Data and Metadata Generation*	♦	♦				CM, MM
IAP	*Integrated Query Processing Schemes*	♦	♦				I, Coll, SB
IAP	*Integrated Metadata*	♦	♦				MM, HH
IAP	*User Characteristics for Personalization*					♦	QP
IAP	*User Modeling for Personalization*		♦			♦	QP
A/V-NTO	*Methodologies for Automatic Multimedia Information Extraction*			♦			CM
A/V-NTO	*Models of Multimedia Content and Context*	♦					CM, MM
A/V-NTO	*Methodologies for Domain and Context Specific Information Extraction*	♦					CM, MM
A/V-NTO	*Multimedia Content Description Standards and Domain Extensions*	♦					HH, CM, MM
UIV	*Visualization*				♦		CM
UIV	*Open User Interface*				♦		SB

Relevance between DELOS Joint Research Activities and BRICKS components

DELOS Cluster	Joint Research Activity	2	5	7	8	40	BRICKS components
ARCH	*Surveying the State of the Art on Network Architectures*		♦				All core bricks
ARCH	*Surveying the State of the Art on Collection level descriptions*	♦					Coll
ARCH	*Development of Peer-to-Peer and Grid Architectures*	♦	♦				All core bricks
IAP	*Development of Conceptual DL Data and Metadata Models*	♦	♦				MM, SB
IAP	*Development of Query Specification and Processing Schemes*	♦	♦				I, SB
IAP	*Development of Metadata Models and Algorithms*	♦					I, MM, SB
IAP	*Development of Data and Metadata Generation Algorithms*		♦				MM
IAP	*Development of Conceptual DL Integrated Metadata Models*	♦	♦				MM, HH

DELOSCl uster	Joint Research Activity	2	5	7	8	40	BRICKS components
IAP	*Development of Integrated Metadata Models and Algorithms*	♦					HH
IAP	*Development of Integrated Query Specification and Processing Schemes*	♦	♦				I, Coll, MM
IAP	*Development of user profiling algorithms*		♦				QP
IAP	*Development of Content Personalization Approaches*		♦				CM
A/V-NTO	*Automatic Annotation Algorithms and Manual Annotation Techniques*	♦		♦			AE, AAG
A/V-NTO	*Content Description Mappings*		♦				HH
A/V-NTO	*Models of Multimedia Databases and Multimedia Peer to Peer computing*	♦					CM
UIV	*Development of a relevance criteria model*				♦		SB
UIV	*Identify the DL users and stakeholders*				♦		All pillars
UIV	*Describe user and usage characteristics*				♦		All pillars
UIV	*Identify user needs, goals and requirements*				♦		All pillars
UIV	*Development of a design methodology and Engineering guidelines*				♦		All pillars
UIV	*User-centred model*				♦		All pillars
UIV	*Mapping user behaviour to user interface*				♦		All pillars
UIV	*Design a dynamic, multilevel, adaptive help facility*				♦		All pillars

BRICKS components

ID	Name	Type	Description
BFC	Bricks Framework and composition	Core	The BRICKS framework is a container for all core Bricks and user Bricks. In addition will it provide user friendly mechanisms for the service composition. The BRICKS framework workpackage will include all tools and libraries for the easy deployment of Bricks according to the BRICKS policy.
SD	Service Discovery	Core	The service discovery service is the most important base service in service oriented architectures. Its function is to discover an appropriate service for certain functionality. Due to its importance to the infrastructure, it will be implemented as a completely decentralized service, so each node is part of the service discovery service.
DMS	Decentralised Metadata Storage	Core	Certain distributed functionalities, e.g. service discovery, require a reliable storage system Hence participants in the BRICKS system must provide storage capacity to the system which forms a decentralized metadata database.
CM	Content Management Brick	Basic	This service will allow users of BRICKS to retrieve contents from every BRICKS node in the network, using a unified service API.
MM	Metadata Manager	Basic	The metadata management provides transformation services, which can transfer a variety of metadata formats to the Metadata Server. BRICKS must offer some (semi)automatic Metadata Extraction services; otherwise

			the initial effort to import large existing archives will be too great for any potential participant.
I	Indexing Brick	Basic	The index Brick will allow BNodes to share index information. Indexes are constructed from local metadata content and used for generating referrals in response to queries. The index should be parametric with respect to the metadata formats and the set of indexed fields.
Coll	Collection Management Brick	Basic	The collection Brick provides a customized view on the content, defined by the user and/or group profile, which allows for specialized searches, etc.
SB	Search and browse Brick	Basic	The Search and Browse Brick will enable users to search documents in a specific collection, or in the default collection including all BNode repositories, according to any of the document dimension, or combination thereof; and to browse collections at different levels of abstraction.
HH	Handling heterogeneous metadata and ontologies	Basic	An important aspect in environments as BRICKS is the integration of different data and metadata sources and ontologies. As it is not possible to define global schemas or ontologies a new approach based on local knowledge of mappings and a distributed learning process between the participants of the system. This process will be part of the ontology Brick.
QP	Query Personalization Brick	Basic	The personalization Brick will cooperate with the Search and Browse Brick to provide personalized-search features to EDM: It will help in reducing the execution time of user queries while increasing (according to user preferences) the recall and precision of queries results.
AE	Annotation Editor	Basic	This Brick is designed for the compilation of information units into coherent contexts by automatically linking related items together. Thus, objects in the user's focus are automatically annotated with not explicitly requested background material or newly-entered items. The service also realises the preparation and publication functionality.
AAG	Access-triggered Annotation Generation	Basic	This Brick is designed for the compilation of information units into coherent contexts by automatically linking related items together. Thus, objects in the user's focus are automatically annotated with not explicitly requested background material or newly-entered items. The service also realises the preparation and publication functionality.
UGA	User/Group Authorisation and Accounting	Core	The authorisation provides the process of identifying what an authenticated user/group can do within the BRICKS system. Accounting enables measuring the resources an authorised user consumes.
DRM	Digital Right Management	Basic	The decentralised Digital Right Management Brick provides protection, and enforces/manage media rights and revenues for the multi-channel digital delivery system.
TM	Trust Model	Basic	Although BRICKS relies on a controlled-environment further a reputation-based trust model is needed to strengthen the overall Bricks trust & confidence infrastructure.
PKI	Decentralised PKI Management	Core	Decentralised public key infrastructure will provide a solution with distributed certification authorities and provides encryption technologies and services that enable secure end-to-end communication.

1.5.6.2   *MinervaPLUS project*

The subject of digitization is clearly of interest to Digital Libraries, and cooperation between DELOS and the MinervaPLUS project has been planned. As a general principle, each project will take the lead and coordinate common activities in the respective priority area: recommendations and guidelines concerning digitisation of cultural contents for MinervaPLUS, and definition of Digital Library architecture, services and interfaces for DELOS.

More specifically, with reference to the workpackages of the MinervaPLUS project, the following cooperation is planned:

- MinervaPLUS WP2: DELOS is invited to collaborate to the development of publications planned by MinervaPLUS. In particular and when applicable, according to the specific topics of the publication, DELOS will prepare a specific chapter were the relevance of that issue will be analysed from the point of view of Digital Libraries.

- MinervaPLUS WP2: the Minerva learning platform will be made available to DELOS, which will propose subjects of common interest for possible joint development of training materials. The DELOS participants will have the opportunity to use and test the training materials developed by MinervaPLUS. DELOS will also facilitate dissemination of the National Policy Profiles on digitisation, developed by the MinervaPLUS project.

- MinervaPLUS WP3, WP4, and WP5: experts selected by DELOS will be invited to join the technical working groups created by the MinervaPLUS project. Guidelines and recommendations developed by these working groups will be made available to DELOS participants, to increase their dissemination and possible application.

- WP6: DELOS and MinervaPLUS will share common tools for identifying good practices and competence centres. In particular the web tools already set-up by the Minerva project will be made available to DELOS project.

- MinervaPLUS WP1, WP7 and WP8: joint DELOS/MinervaPLUS coordination meetings are expected to be organised in conjunction with the plenary meetings of the NRG (every six months). The joint DELOS/MinervaPLUS meetings will aim at monitoring the status of advancement of the joint activities and at preparing the next joint initiatives for the following six months.

Symmetrically, with reference to the workpackages of DELOS, the following cooperation is planned:

- DELOS WP1: MinervaPLUS will be invited to contribute to the user requirements for the DELOS DL reference architecture. The reference architecture will be disseminated to the MinervaPLUS community through a workshop organized by DELOS. DELOS and MinervaPLUS will jointly disseminate the reference architecture to the large communities of Librarians, Archivists and Museums.

- DELOS WP2, WP3, and WP5: DELOS will disseminate to the MinervaPLUS community the results of the activities carried out by WP2, WP3, and WP5 in the areas of information access, information personalization, metadata extraction, and semantic interoperability. Joint awareness events will be organized to disseminate widely these results.

- DELOS WP4: MinervaPLUS will be invited to provide user requirements for the design of user interfaces and visualization techniques for digital libraries. DELOS will disseminate the results of this activity to the MinervaPLUS community.

- DELOS WP6: Joint DELOS-MinervaPLUS awareness activities will be promoted aiming at raising the profile of digital preservation issues within the digital library community.

- DELOS WP8: A coordination meeting will be organized to define a joint plan of dissemination activities. Particular attention will be paid to the following DELOS dissemination activities: (i) summer school, (ii) brainstorming workshop, (iii) awareness events, for which collaboration with MinervaPLUS is highly desirable.

- In addition, representatives from DELOS will join the MinervaPLUS working groups to contribute with their expertise and to bring the digital library perspective within them.

1.5.7 International Cooperation

One of the DELOS network's objectives is to capitalize on leading European expertise to become an international center of excellence, acting as a reference point concerning Digital Library initiatives. Therefore, focus will be to involve also non-EU partners in DELOS activities to bring an international dimension to the Network.

The DELOS network will thus seek to establish collaborative actions with other research communities. In particular, collaborations already established by the FP5 DELOS NoE (with the US, Russian, Central and Eastern European, and non-EU Mediterranean Digital Library research communities) will be reinforced and extended to include other countries (Latin America, Japan, China, India) with which an agreement for scientific cooperation under the 6th FP is defined by the European Union.

DELOS-US
The past collaboration between the FP5 DELOS and the National Science Foundation (NSF) took the form of seven joint Working Groups (EU-US) on the topics of: Spoken-Word Digital Audio Collections, Digital Library Information-Technology Infrastructures, Personalization and Recommender Systems in Digital Libraries, Emerging Language Technologies and the Rediscovery of the Past, Digital Imagery for Significant Cultural and Historical Materials, Preservation and Archiving, Agents in Digital Libraries and the organization of a series of jointly sponsored thematic workshops. This collaboration has been very successful and, thus, the DELOS network intends to continue in the same direction, improving and extending this collaboration. In particular, the DELOS network will promote the establishment of joint working groups in key technical areas. These groups will foster an international community of Digital Library researchers and will focus on generating joint research agendas. This collaboration will, hopefully, lead to joint research activities and projects.

DELOS-Russia
A Digital Library Programme of the Russian Foundation for Basic Research, which involves 20 projects is now underway. The DELOS network intends to establish a cooperation agreement with this Russian initiative. The objective of this collaboration will be to promote the exchange of results and experiences between the European and Russian Digital Library communities, to disseminate information on Digital Library research and development carried out in the EU to the Russian community and viceversa, to provide for participation of the Russian Digital Library research community in collaborative working groups and projects, and for the organization of joint technical and scientific events. The past collaboration between the FP5 DELOS and the Russian Digital Library Programme had taken the form of a jointly sponsored thematic workshop (Moscow June 2001).

DELOS-Latin America
The ISTEC - Ibero-America Science and Technology Education Consortium - has expressed formally its willingness to be associated with DELOS in a future possible DELOS-FP6. It is also submitting a similar proposal to the NSF. This collaboration would focus on (i) Training and Technology transfers: participation of ISTEC members in the Digital Library Summer School, supported by DELOS; possible participation of DELOS members in similar actions promoted by ISTEC; common support of ISTEC and DELOS to training actions. (ii) Workshops and working groups: creation and participation of ISTEC and DELOS members in international working groups, similar to the DELOS-NSF working groups. (iii) Conferences: cooperation between ECDL- European Conference on Digital Libraries, supported by DELOS, and similar events promoted by ISTEC.

DELOS-Central and Eastern European Countries
The FP5 DELOS NoE has promoted its activities and achievements to the Central and Eastern European countries through an extension of its contract funded by the EU. This cooperation had

taken the form of thematic workshops and training and tutorial activities and had involved a number of universities and cultural heritage institutions of these countries. The DELOS network intends to continue in the same direction by improving and extending this collaboration.

DELOS-Non-EU Mediterranean Countries

The FP5 DELOS NoE has established collaboration with the non-EU Mediterranean Countries Digital Library research communities. This collaboration took the form of a jointly sponsored thematic workshop (Morocco November 2001) and the participation of members of these communities in the DELOS Summer School. The DELOS network intends to continue in the same direction by improving and extending this collaboration.

DELOS-China, DELOS-India and DELOS-South Africa

Some of the institutions participating in DELOS have already established scientific links with China, India and South Africa. With the assistance of those participants, the DELOS Network intends to explore the possibility of establishing closer links and to propose joint events and some form of cooperation.

1.6 Gender Action Plan

There are no particular gender issues arising from DELOS, apart from those that have been affecting the European research community from the beginning. Gender parity is still far from being achieved in research activities in the Information Technology sector, where most of the graduates come from scientific disciplines (especially Engineering), which are still predominantly male. Indeed, female representation in Information Technology (engineering, mathematics and computer science) reaches a humble 27,6%, far behind other academic activities, such as arts, literature and even medical / biological studies. The European Commission and dedicated gender equality institutions have defined a common framework to promote parity, and DELOS will support the existing parity strategies established by expert committees in Europe. In particular, DELOS is committed to encouraging and supporting the participation of women in research and in ensuring that research activities reflect and address the needs of women as well as men. In this perspective, the joint programme of activities will be implemented with the following underlying gender equality supporting measures:

With respect to supporting the participation of women in research, the set of activities where the Network can have the most effective impact in striving for gender equality is in the education and training programmes, such as tutorials, summer schools, etc. Efforts will be made to increase the participation of female researchers in those activities through dedicated grants and scholarships with the objectives of achieving eventually at least a 40% female participation. In addition, attendance of female researchers at workshops and other scientific events will be supported again through dedicated grants. The Network will also do its best to stimulate vocations, create incentives, support female researchers in pursuing their careers and to promote research as a real opportunity for younger female students.

The Network will focus on a variety of contextual factors that have a role to play in the development and use of Digital Library systems. Studies designed to investigate these factors will also collect gender-based data on participants – in particular in the user requirements surveys included in the activity plans of two of the Network's clusters: Information Access and Personalisation, and User Interfaces - and will include gender as a factor in experimental analysis.

The Network will aim to reach a 40% female representation not only in the education and training activities, but also within the different bodies and structures involved in all the other activities of the Network. Several gender equality committees, workshops and symposiums, have identified the figure of 40% as a major long term parity objective in the male-driven Information Technology domain.

Within the scope of DELOS, focus will also be on fighting discrimination affecting women. Indeed, discrimination is the major cause of women dropping out of research. Several studies have highlighted the fact that women still suffer harder selection processes and reduced wages (up to

25%) than their male counterparts. As such, women have developed a broader range of skills to make up for their status.

1.7 Raising public participation and awareness

In order to reach a wider audience and improve the involvement of different national application communities beyond the research community, the members of the Network will establish contacts with national application communities (i.e., libraries, archives, museums) for the purpose of organizing joint workshops on themes of mutual interest. These types of workshops will allow the Network to raise public participation, help spread awareness and explore the wider societal implications of the proposed work on digital libraries.

A knowledge transfer programme will also be organized which will consist of specific knowledge transfer courses in selected application domains. These courses will present digital technologies and successful case studies in a customized manner to user communities. They will be focused on specific application environments, on their demonstrative mission, and on their window of opportunity.

These knowledge transfer actions which will be undertaken by DELOS NoE aim at convincing European user communities of the effectiveness and benefits of the employement of digital library technologies within application/commercial environments.

It is widely accepted that the lack of a quick and wide-spread *"take-up"* of new technologies is one of the major problems of Europe, compared with other continents. The DELOS NoE will disseminate information on its activities not only to the research world, but also to the relevant application and industrial communities and will undertake the following take-up measures:

International Summer School.
The Summer School will provide a course on the emerging domain of Digital Libraries and its underlying technologies. Moreover, it will offer a unique opportunity for a complete update on all aspects of DLs and will encourage contacts between participants, lecturers, and interested parties.

Training
A technology transfer programme will be run consisting of a series of events to be organized in collaboration with specific application and industrial communities. The aim of these events is to present a new DL technology or a new system to a particular community and demonstrate how it could improve the services provided or the productivity - depending on the type of community involved.

1.8 Major Milestones over full Network duration

As stated in the first section, the major objective of DELOS is to help the integration and the coordination of the ongoing research activities of the major European teams working in DL-related areas, with the goal of developing the next generation DL technologies. The Network aims at developing generic DL technology to be incorporated into industrial-strength DL Management Systems (DLMSs), offering advanced functionality through reliable and extensible services.

The first year of activity will be mostly devoted to bringing together the research teams and to perform extensive exchange of ideas and experiences. In the second year the main technologies needed in a Digital Library Management System will be defined and assessed. In the third year a reference architecture for a generic DLMS will be defined. In the fourth year DLMS prototype (possibly with limited services and functionality) will be tested and evaluated in some relevant application areas. In summary, the major milestones for the DELOS Network are the following.

M1 month 12	*DELOS Web site and Portal.* The DELOS web site will maintain and make accessible the collection of all the results and reports made available by the Network and its participants. It will also provide information on the latest research results in the field of digital libraries as well as the latest information about international projects, initiatives, conferences, etc in the digital library domain, and provide access to all major Web sites related to digital libraries. In addition, the web site will provide tools to facilitate the exchange of information and the discussion of research topics among the network participants
M2 month 24	*Cross-referenced survey of all DL-related fields.* An extensive survey of the technologies and the state of the art in all the DL-related fields will be carried out and made available to the research community. The different technologies will be examined and cross-referenced, in order to provide indications for the main components of a DLMS.
M3 month 36	*Reference architecture of a Digital Library Management System (DLMS)* A reference architecture for a generic DLMS will jointly defined by the Network, based on the concepts and indications provided in the survey. The main components of the DLMS will be identified and defined, after having determined the most suitable technologies available in the different DL-related fields
M4 month 48	*Limited-functionality DLMS prototype and joint evaluation* An implementation of the new architecture will be carried out jointly by some of the Network participants. A limited-functionality DLMS prototype will be tested and evaluated in some specific application area, in order to identify the strengths and weaknesses of the reference architecture. The results and the refinements of the demonstrator system will be made available to the research community.