

SIXTH FRAMEWORK PROGRAMME

PRIORITY IST-2002-2.3.1.12

Technology-enhanced Learning and Access to Cultural Heritage



Information Society
Technologies



Contract for:

NETWORK OF EXCELLENCE

Annex 1 - "Description of Work"

Detailed JPA for the period January 2005 – June 2006

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 original Annex 1: **01 December 2003 (FINAL)**

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

Detailed joint programme of activities 2 (JPA2) – (month 13 to 30)

Introduction

The DELOS Network of Excellence is carrying out a broad range of interrelated activities whose combined effect should be to contribute to the successful achievement of the objectives as defined in Section 2 of the Technical Annex.

During the first year, DELOS activities focused on tightening up the Network by dedicating efforts to constructing a strong community sense among the Network members, and setting up the necessary infrastructure (Website, DELOS Digital Library, BSCW) to support collaboration and integration between the members. The Joint Programme of Activities for the second period (JPA2) has been defined as a result of a Call for proposals among the DELOS members. At the end of an exhaustive selection process, the Scientific Board selected 25 of the 59 proposals received. The goal of JPA2 is to achieve more concrete results (demonstrators, prototypes) while continuing to pursue the following important objectives of DELOS.

Further development of Digital Library enabling technologies. JPA 2 will concentrate on the development of those technologies that enable the building of the next generation digital library management systems.

Improvement of cross-cluster integration. One of the aims of JPA2 is to improve the integration of research throughout the Network. Joint research activities by partners belonging to different clusters has thus been encouraged and most of the activities of the JPA2 will be conducted by partners belonging to several clusters.

Wide dissemination of the DELOS results. A wide dissemination activity directed in particular towards the interested application communities will be undertaken in order to disseminate the results achieved by the Network during the first and second year of activities.

The activities of JPA2 will continue to be organized into topic-specific clusters. Each cluster is structured according to the Workpackage-Task hierarchy. In addition, several workpackages cover aspects regarding DELOS management and dissemination. The Network is thus composed of the following Workpackages for the second 18 month period:

- WP1 Digital Library Architecture
- WP2 Information Access and Personalization
- WP3 Audio/Visual and Non-traditional Objects
- WP4 User Interfaces and Visualization
- WP5 Knowledge Extraction and Semantic Interoperability
- WP6 Preservation
- WP7 Evaluation
- WP8 Dissemination and Spreading of Excellence
- WP9 Assessment
- WP10 Administrative Management
- WP11 Scientific Management

This addendum to the TA provides a brief description of the activities to be conducted by the Network during the second 18-month period. More complete information can be found in the following documents: “JPA2_Detailed_Description.doc” and “JPA2_Selected_Proposals.doc”.

WP1 - Digital Library Architecture

The overall goal of this workpackage is to analyze, develop, and integrate architectures for digital libraries.

Strategic Objectives

During the next 18 months (January 2005 – June 2006) the primary goals of the first 18 months of the project will continue to be actively pursued, with the last two topics being addressed in more detail during the first six months of 2005.

- Developing new approaches to the architecture for an “intelligent management” of digital libraries
- Enabling the coordinated development of information architectures by an adoption of a set of common standards and protocols
- Managing information dynamics and mobility

Workpackage Activities and Integration

General Digital Library Architecture issues will continue to be considered in JPA2, with the main focus aiming at the development of a DL reference model. At the same time, two new tasks in JPA2 will more concentrate on digital library architectures for special purposes. The main focus will be on Digital Library Architecture for e-Health applications. One of the goals of WP1 will therefore be the investigation and provision of a dependable platform. Stream data play an important role in e-Health applications. So, the platform must also support stream processing based on the integration of stream operators and web services. In addition to the tasks already described in JPA1, the following JPA2 tasks will be pursued.

A Reference Model for Digital Library Management Systems.

The objective of this Task is to introduce a reference model for Digital Library Management Systems (DLMS), i.e. a formal and conceptual framework describing the characteristics of these particular kind of information systems. Task activities will address the following issues:

- Evaluation and survey on architectural frameworks for digital libraries.
- Current digital library systems: user requirements vs provided functionality.
- User requirements.
- Survey on Current Digital Library Systems and Gap Analysis.
- Definition of a Reference Model for DLMSs.

Design, Implementation and Evaluation of Multimedia Annotations for Users' Collaboration.

The main goal of this Task is the implementation and the evaluation of an annotation digital library service, based on the design guidelines and on the existing partners available tools. In particular, task activities will address the following issues:

Design and Implementation of the annotation digital library service. The most important requirements that should be fulfilled by the annotation service are the following:

- Support of nested annotations, i.e. not only documents or document parts can be annotated, but also other annotations can be annotated.
- Possibility to reference each annotation by an handle (e.g. the Uniform Resource Identifier (URI) could be one of the schemes to be supported).
- Faithfully represent the sign (e.g., textual, graphical, referential or a combination of these) and the meaning of an annotation, so that different annotation types can be supported.
- Support different scopes of annotations (e.g. private, public, shared).

Definition of a set of APIs to allow the access to this service from different digital libraries. The overall goal of this task will be accomplished by integrating also contributions from WP4 and WP7.

Management of and Access to Virtual Electronic Health Records.

This Task is the logical continuation of T1.1 of JPA1. It evaluates the architecture and adds in it aspects of electronic health records, which represent an important application field for digital libraries. The realization of these goals requires an infrastructure that is highly dependable and reliable. Moreover, the infrastructure has to allow for the transparent access to distributed data, and to efficiently schedule the access to computationally intensive services by applying sophisticated load balancing strategies using GRID technology. Common standards as investigated in JPA1, task 2 are equally important. Task activities will address the following issues:

- Identification of the basic building blocks to access distributed artifacts and to intelligently search within a set of these artifacts.
- Provision of a dependable platform that supports the integration of these building blocks into processes (e.g., based on the ETH/UMIT hyper database prototype system OSIRIS – Open Service Infrastructure for Reliable and Integrated process Support), thereby realizing a virtual electronic patient record.

Integration of Data Stream Management into an eHealth Digital Library.

This task is a continuation of Tasks 1.1 and 1.3 of JPA1, considering socially relevant applications of future digital libraries to e-health and health monitoring. Continuous data streams generated by (wearable) sensors have to be processed online in order to detect critical situations. In addition to the stream operators, also traditional discrete (web) services, e.g., services that do not operate on continuous input data, have to be integrated. Task activities will address the following issues:

- Survey on use cases from tele-monitoring applications containing the combination of stream operators and web services.
- Specification and implementation of join and search operators for data streams.

Expected Results

Reference Architecture: The reference that will result from the Task described above will strongly contribute to enhance and improve the quality of the DLMSs since it will specify their expected features and properties. It will also lay the foundations for establishing what we have achieved until now, where we want to move in the future, what we should do and how we can evaluate priorities and measure advances.

Design, Implementation and Evaluation of Multimedia Annotations for Users' Collaboration: Implementation of an annotation digital library service, based on the design guidelines and on the existing partners available tools, and defining a set of API to allow the access to this service from different digital libraries. Integrating this service into the DAFFODIL and BRICKS digital library systems.

Management of and Access to Virtual Electronic Health Records Implementation of sample building blocks and processes in combination with the specialized applications of HITT/TILAK

Integration of Data Stream Management: Prototype implementation of an infrastructure for Workflow processes including stream processing supporting the integration of stream operators and web services.

WP2 - Information Access and Personalization

The main activities of WP2 are organized along three dimensions: access of information in a single provider; integration of information in multiple providers; personalization, i.e., customization of system behavior to the user.

Strategic Objectives

During the next 18 months (January 2005 – June 2006) the Information Access and Personalization Cluster will pursue the following strategic goals:

- Support of cooperation among individual research groups that will promote common consensus among researchers.
- Research on new models and methodologies in order to overcome inefficiencies of existing ones.
- Development of toolkits or systems for purposes of re-use and demonstration of proposed methods and models.

In addition, the main goals of the first year of the project will continue to be actively pursued:

- Promotion of knowledge about available practices in the fields of information access and personalization in digital libraries.
- Construction of a common, comprehensive framework for information access and personalization approaches.
- Initiation of research on new information access and personalization models and methodologies.

Workpackage Activities and Integration

Current plans towards all these goals come both as a product as well as a logical continuation of the work of the cluster so far. During the first year, in addition to information sharing and dissemination activities that have taken place (Summer School, cluster portal, etc.), the cluster mostly concentrated on drafting a set of comprehensive surveys and reports on some of the key relevant areas of interest, to provide broad overviews of existing models and approaches and outline problems. These surveys formed the basis for both establishing common foundations on information access, information integration, and personalization (Tasks 2.1-2.3, respectively) and initiating joint research on some of these topics (Task 2.4 for information access and Task 2.5 for personalization), becoming the reason for several researchers' exchanges. In particular, the surveys drafted are on the following topics: "Information Access Models and Modes", "Metadata in the Context of DL", "Foundations for Information Integration", "Peer-to-Peer Data Management Systems", "Semantic Interoperability in Digital Libraries", "Data Annotation and Provenance in Large Scale Information Integration Systems", "Data Profiling for Personalization in Digital Libraries", and "Personalization Methods for Content Selection". Two of these surveys are completed and available through the cluster website, while the others are expected to be completed in the first six months of JPA2. These surveys will be combined into a single volume to be delivered after the completion of all.

Furthermore, several interesting and significant problems have been selected as the next research agenda of the cluster, organized within Tasks 2.6-2.10, outlined below. These can be considered as specializations of Tasks 2.1-2.5, which in their current form will remain active for the first six months of JPA2. During that period, as part of these tasks, in collaboration with the "Architecture" cluster, a workshop titled "Future Digital Library Management Systems" will be held in Dagstuhl, Germany, March 29th-April 1st, 2005. In addition, as part of these tasks, initial toolkit specifications for data search algorithms and query personalization algorithms will be delivered. (The one on personalization algorithms (D.2.6.2) will be delivered on month 18 instead of 12, as the latter was a typo in the original JPA.) The second workshop planned for JPA1, titled "Personalization in Digital Libraries", originally planned for the summer of 2005 has been postponed for the summer of 2006 (as part of Task 2.0). The reason for the change is to avoid conflict with the "Personalized Information Access" workshop that has been independently initiated as part of the "User Modelling" Conference, in Edinburgh, UK, July 23rd-19th, 2005. This event will now be co-sponsored and co-organized by DELOS, hence maximizing the breadth of the relevant networking activities. With the conclusion of the more general tasks on foundations and research initiation (i.e., Tasks 2.1-2.5), Tasks 2.6-2.10 will become the sole activities of the cluster. Brief descriptions of these are given below.

Task 2.6 - Advanced Access Structures for Complex Similarity Measures.

The main objective of this task is to broaden the filtering approach towards metric indexing in order to provide indexing proposals for metric spaces that (i) efficiently support nearest neighbour, range, and ranking queries, (ii) operate on any kind of multimedia data, and (iii) are generic in the distance measure to be employed. This task is a continuation of the activities of Tasks 2.1 and 2.4 from the previous JPA. Its overall goal will be accomplished by integrating also contributions from WP3.

Task 2.7 - Application of the P2P Paradigm in Digital Libraries.

The aim of this task is to investigate the peer-to-peer (P2P) resource-sharing paradigm for large-scale distributed Digital Libraries (DL). The objective is to support decentralized sharing of data and services in a network of autonomous and heterogeneous DL nodes. Two fundamental aspects of the P2P paradigm in Digital Libraries (P2P DLs) will be considered: (a) query reformulation, and (b) query roaming in P2P DLs. This task is a continuation of the activities of Tasks 2.1, 2.2, and 2.4 from the previous JPA. Its overall goal will be accomplished by integrating also contributions from WP1

Task 2.8 - Personalized Query Routing in Peer-to-Peer Federations of Digital Libraries.

This task explores routing of various types of queries (SQL, XQuery, etc.) over a P2P network where, apart from DLs, user agents with powerful personalized tools may participate as peers as well. It attempts to design novel strategies for query routing that can effectively exploit user profiling information about the participating users. This task is an offspring of all technical tasks (Tasks 2.1-2.5) from the previous JPA. Its overall goal will be accomplished by integrating also contributions from WP1.

Task 2.9 - Context-dependent Access to Digital Libraries.

This task investigates how the notion of context can be incorporated into data stored in DLs of relational form. The investigation will be done at two levels: an abstract and an applied one. Initially the focus will be on laying the foundations for a context-aware database management system, and then the task will attempt to define a uniform mechanism for personalized access to Digital Libraries. This task is a continuation of the activities of Tasks 2.1, 2.3, 2.4, and 2.5 from the previous JPA. Its overall goal will be accomplished by integrating also contributions from WP3.

Task 2.10 - Modelling of User Preferences in Digital Libraries.

The aim of this task is to study the provision of services for the management of user preferences, focusing on query personalization (enhancing a query by incorporating into it user-specific preferences), user notification (notifying a user when an “event” of interest happens), and document customization (allowing the user to compose documents that will be “materialized” by the system). This task continues Tasks 2.3 and 2.5 from the previous JPA.

Expected Results

Through the work performed on the above topics as well as the composition of the surveys, research advances on the following major themes in information access and personalization are expected:

- Information access: data indexing for complex similarity measures
- Information integration: query processing and routing in P2P architectures
- Personalization: modelling of user preferences and more general contexts

WP3 - Audio/Visual and Non-traditional Objects

The long term research activities for WP3 have as targets Metadata Capturing for Audio-Visual Content, Universal Access and Interactions with Audio-Visual Libraries, and Management of the Audio-visual Content.

Strategic Objectives

The new tasks of WP3 fit well with the above three targets. In particular the new six tasks listed under WP3, along with other tasks (listed below) in which WP3 members participate (and which are partially overlapping with the above WP3 objectives) cover the three main objectives of the WP3 as follows:

Metadata Capturing for Audio-Visual Content:

- Video Annotation with Pictorially Enriched Ontologies (WP3 new task)
- Automatic, Context-of-Capture Based Categorization, Structure Detection and Segmentation of News Telecasts (WP3 new task)
- Multimedia Interfaces for Mobile Applications (WP3 new task, in cooperation with WP4)

Universal Access and Interactions with Audio-Visual Libraries:

- Content and Context-Aware Multimedia Content Retrieval, Delivery, and Presentation (WP3 new task)

- Description, Matching and Retrieval by Content of 3D Objects (WP3 new task)
- Natural Language and Speech Interfaces to Knowledge Repositories (WP3 new task)

Management of the Audio-Visual Content:

- Advanced Access Structures for Similarity Measures (WP2 task, in cooperation with WP3)
- Interoperability of e-Learning Applications with Digital Libraries (WP2 task, in cooperation with WP5)
- Ontology Driven Interoperability (WP2 task, in cooperation with WP5)

Workpackage Activities and Integration Activities

JPA1 - During the first 12 months (JPA1) the work in WP3 aimed to develop a common understanding and a foundation of the work that has to be done in DELOS in terms of State of the Art reports, support for Forum and Testbeds, and efforts at understanding the expertise of the partners and their possible cooperation towards the DELOS objectives as they are described in the Technical Annex. In particular we had a large number of integration activities which among others included the production of three extensive State of the Art reports with the collaboration of all the cluster members in three areas that cover the research focus of the audiovisual and non-traditional objects cluster and related topics. The integration activities of the first year also included the mapping of expertise and the potential contributions of the members of the cluster to Delos objectives, as well as extensive collaboration to define joint research proposals with the participation of member organizations from other clusters. Several of those joint integration and research activities were approved for financing by Delos in JPA2.

JPA2 - The efforts of the next 18 months (JPA2) will build on the established infrastructure, experiences, and cooperation that have been established in JPA1. The integration activities already activated in JPA1 will continue and expand in JPA2. Cooperation between WP3 cluster partners and with partners of other clusters will focus on specific objectives as defined in the projects approved, that cover some of the strategic subjects of investigation of future audio visual digital libraries. The technical and scientific goals and objectives of the approved projects are described in more detail in the JPA2 Workpackage Description. For each project, partners have been selected so as to provide complementary expertise. Projects will allow focused exchange of know how so as to enlarge the individual partner's knowledge base, partners' cooperation in order to advance the current state of the art on specific subjects of investigation, joint cooperation for the development of prototype systems that provide proof of concepts.

Cooperation with other DELOS clusters will be improved since most of the projects are cross-clusters, including research groups that in JPA1 belonged to other clusters. On the other hand, researchers of WP3 will also participate into projects coordinated by other clusters, as described in JPA2 Workpackage Description.

Projects will allow research exchange on clearly identified subjects of joint investigation. They will also permit exchange between institutions of software modules individually developed.

Testbeds and demonstrators software will be finalized, and a number of new demonstrators, including projects' outcomes will be added.

Together with WP4 a Workshop will be organized in Cortona, Italy, on May 4-6, on "Audio Visual Content and Information Visualization on Digital Libraries" that aims at gathering researchers on user interfaces (WP4) and on audio visual digital libraries (WP3). In the workshop we plan to organize a panel on "Applications of Digital Libraries to Cultural Heritage and Related Issues" that will bring together representatives of the major current EC projects and initiatives like EVA, BRICKS, EPOCH, and MINERVA to plan cooperations and synchronization of activities and results.

Cluster meetings will be carried out on a periodical 6-month basis. The first JPA2 cluster meeting is planned to take place in Cortona, the day before the workshop

In particular the following joint research and integration activities have been defined for JPA2 under WP3:

- *Video Annotation with Pictorially Enriched Ontologies*: defining methodologies and techniques to describe concepts and their specializations by augmenting an ontology of linguistic terms with "visual concepts"

- *Automatic, Context-of-Capture Based, Categorization, Structure Detection and Segmentation of News:* investigation of methodologies for the automatic classification of news stories utilizing visual and audio features
- *Multimedia Interfaces for Mobile Applications:* construction of video summaries upon user's request, and design of interaction for mobile devices
- *Description, Matching, and Retrieval By Content of 3D Objects:* supporting structural and view based retrieval by content of 3D objects by similarity.
- *Content and Context Aware Multimedia Content Retrieval, Delivery and Presentation:* providing intelligent retrieval in large media collections with visualization adapted to user preferences.
- *Natural Language and Speech Interfaces to Knowledge Repositories:* providing principles, methodologies and software for the automation of the construction of natural language and speech interfaces to knowledge repositories.

In addition, as mentioned above, WP3 member organizations are participating with larger or smaller participation in a number of other tasks which are listed under different workpackages thus contributing to cross-cluster integration in Delos.

Expected Results

We expect that during the JPA2 the integration of the cluster members will be advanced in several ways and in particular that it will become more focused through the collaboration of partners in subprojects (tasks) with specific objectives.

We expect that a number of prototypes (as described in the respective tasks) will be developed and that they will be used as demonstrators of the Delos common understanding and State of the Art. The software that is currently been developed for supporting testbeds and demonstrators will be ready and will provide uniform access to those prototypes.

A number of research results will be derived in each of the specific subjects of collaboration and joint research as described in the respective tasks. Those research results will lead to joint publications of the Delos members, and, equally important, to new joint perspectives of the State of the Art and the research needs among Delos partners.

Finally, the research collaborations, the prototype demonstrators, and the interaction of the Delos members in Delos workshops and cluster meetings will lead to the identification and definition of new research activities and collaboration among Delos partners.

WP4 - User Interfaces and Visualization

The ultimate goal of the User-Interface and Visualization cluster is to develop methodologies, techniques and tools to establish a theoretically motivated and empirically supported frame of reference for designers and researchers in the field of user interfaces and visualization techniques for digital libraries, so to enable future DL designers and developers to meet not only the technological, but also the user-oriented requirements in a balanced way.

Strategic Objectives

Specific objectives of the WP are:

- To elaborate a common understanding of the role and scope of user interface research in the digital library area.
- To develop a theoretical framework for digital library user interface design.
- To develop user-centered methodologies, techniques and tools to be exploited by DL designers and developers.

During JPA1 the research concentrated mainly on the understanding of the role and scope of DL user interface research and the investigation of the requirements for a DL interface design that provides support to the user throughout the entire DL lifecycle. A number of sources of information have been studied for this purpose, they include: relevant research efforts/projects in the DL domain; DLs users,

including stakeholders such as maintainers, librarians, and archivists; DL literature. The overall analysis of these sources has been reported in deliverable D4.1.1. Based on this foundation, in JPA2 the cluster will address the development of the above-mentioned theoretical framework, methodologies and tools.

Workpackage Activities and Integration

During JPA1 many collaborations among members of the Work Package have been initiated, including two PhD student exchanges supported by Delos, giving rise to a strong collaboration and integration among members both in the production of the deliverable and in the organization of the Delos workshop on Audio-Visual Content and Information Visualization in Digital Libraries (AVIVDiLib'05), held in Cortona on May 4-6, 2005, jointly with WP3. During JPA2 the intercluster collaboration will be much enforced, since all but one of the tasks are conducted in cooperation with other clusters, as specified in the following.

Task 4.5 – Visualization in DL systems – aims at providing tools and techniques exploiting visual clues in making sense out of information, focusing on two specific application scenarios: content-based multimedia retrieval and small screen interfaces. For WP4, it is a continuation and evolution from JPA1, but it exhibits a much strong integration since it now also involves contributions from WP1, WP2 and WP3.

Task 4.7 - User Requirements-driven Support for a DL Design Framework - builds over the results of Tasks 2 and 3 of JPA1 and aims at extending such results towards the systematic investigation of non-conventional interaction paradigms, and the correlation of such paradigms with different usage phases of digital libraries. Its final aim is to define (and implement) advanced interaction paradigms for DL and thereby build a theoretical framework for the design of new DL interfaces.

Task 4.8 - Task-centered Information Management - builds over the results of Tasks 3 and 4 of JPA1 and aims at realizing another part of the envisioned theoretical design framework, through the task-oriented modeling of user interaction in digital libraries, with the final goal of realizing a system able to learn the way in which users carry out typical tasks and then almost automatically reproduce them. The overall goal of this task will be accomplished by also integrating contributions from WP2.

Task 4.9 - Natural Language and Speech Interfaces to Knowledge Repositories - this task is in strict cooperation with WP3. As for WP4, it represents a continuation of Task 4, since context and accessibility considerations pushed WP4 people to further investigate interaction paradigms other than visual ones in particular concentrating on those exploiting the auditory channel. Its main objective is to provide principles, methodologies and software for the automation of the construction of natural language and speech interfaces to digital libraries.

Task 4.10 - Design, Implementation and Evaluation of Multimedia Annotations for Users' Collaboration - this task continues the already very successful collaboration among the involved DELOS partners on a comprehensive annotation model and web annotation system which will serve as design guidelines for the specification and implementation of an annotation service compliant with contemporary interface standards. The overall goal will be accomplished by integrating also contributions from WP1 and WP7.

Expected Results

More specifically, the major expected results for JPA2 are:

- Further collection and analysis of end-users' and stakeholders' requirements. This would facilitate a richer understanding of the causal nature that exists between the respondent, functional requirements, as well as the non-functional requirements and would make it possible to use the results to refine the existing lifecycle model.
- Definition of a design framework based on the revised lifecycle model and equipped with concrete examples of suitable interaction paradigms, both visual and non-visual, e.g., natural-language based. Such a design framework will also consider more advanced task-oriented interaction mechanisms and semantic views of the available information.

- Actual implementation and evaluation with real users of specific tools representing valuable instantiations of some of the advanced functionalities covered in the theoretical framework such as collaboration support, automatic linking, recommendation, visual exploration and enrichment.

WP5 - Knowledge Extraction & Semantic Interoperability

The cluster has established a common foundation and understanding of the state-of-the-art in two of the three designated task areas during the first twelve months of the Network. Work on the third task area (digital repositories) will be completed during the next six months as scheduled.

Strategic Objectives

Over the next 18 months we will pursue the following strategic goals:

- To complete foundation work carried out during the first 12 months, describing the state-of-the art in the three task areas of information repositories, knowledge extraction and semantic interoperability
- To develop prototypes which demonstrate the application of data models, metadata standards, metadata schema profiles and schema mappings for enhanced semantic interoperability
- To work in the domains of e-Learning and cultural heritage to inform the user requirements, design, testing and evaluation of the prototype systems.

Workpackage Activities and Integration

Work in the task area of Information Repositories will continue jointly with the Preservation cluster (WP6), however, there are also major investments in e-learning repositories, for example in the UK, the JISC Online Repository for Learning and Teaching Materials (JORUM) has a national role. In addition, there is increasing implementation of institutional repositories, which are providing a secure and trusted open archive for learning objects. These applications have highlighted the importance of metadata standards in this context. Many of the learning objects are complex objects and are presented as packages, which include multimedia items offering richness to the teaching process. There is synergy with the Audio-visual cluster (WP3) in this area and a User Evaluation (WP7) requirement.

The study on semantic interoperability clearly shows the need for more work to investigate the relationship between such systems to explore more specific integration issues related to the mapping of a high-level ontology to a domain-specific ontology, and to demonstrate these complex mappings in a working prototype. In order to test the mappings to a domain ontology it is necessary to work closely with a discrete community of users and the cultural heritage sector has been selected as providing a rich user base and one where the partners have existing expertise and experience. This research will be addressed through a new task, which also draws in expertise from partners in WP2 and WP7.

Task 5.1 Information Repositories & Open Archives

This Task activity aims to bring together repository developers, self-archiving members of faculty and librarians in a multidisciplinary Forum which will have an electronic presence but will also be supported by a workshop and report. The overall goal of this task will be accomplished by integrating preservation aspects of repositories through contributions from WP6.

Task 5.4 Interoperability of eLearning applications with digital libraries

This Task will concentrate on the interoperability of audiovisual digital libraries and eLearning applications, to support the modular development of interactive personalized learning experiences. In particular, the work will focus on semantic mappings between learning standards and audiovisual content standards leading to development of a robust model that will facilitate use of semantic descriptions of audiovisual content objects, in the creation of reusable learning objects.

Task 5.5 Ontology-driven Interoperability

The objective of this Task is the investigation and development of methods for the integration of heterogeneous data types, models, upper level ontologies and domain specific knowledge organization systems (KOS). This effort will be realised via research reports, guidelines, real world case studies and

a pilot development demonstrator. The experimental material is taken from the rich cultural heritage domain and traditional library science.

Expected results

The major expected results for JPA2 are:

- The Digital Repositories Forum which will support discussion of particular topics contributing to an evaluative study on the development and implementation of community repositories to support research (institutional, national, e-prints, subject/disciplinary, species, scientific and other e-data) and learning & teaching (institutional, national learning objects and other materials). The study will be informed by a Thematic Workshop.
- The major standards for eLearning (e.g. SCORM) and A/V content description (e.g. MPEG7 etc.) will be comparatively studied and a mapping will be developed to describe both the complementarities of standards and the alternative use of the standards. Based on this analysis, an integration framework will be developed and additional mappings will be defined using XML. An appropriate digital library management component demonstrator will be implemented and the response of different classes of users (digital library users and distance learners will be assessed.
- Semantic interoperability goals include participation in the CIDOC CRM-SIG and IFLA-FRBR Review Group for the harmonization of FRBR and CIDOC CRM. The Core-Ontology to Schema mapping work will include defining structures for documentation units and structures for simplified querying and a mapping to TEI. There will also be a mapping of the CRM to Dublin Core (DC) to enable automatic DC access to CIDOC CRM compatible resources. There will be a review of facet analysis in Library & Information Science and Semantic Web/Grid communities. Lastly, a proof of concept demonstrator for KOS mapping will be built and evaluated to assess its application value and to provide a view in relation to other ontology management tools.

WP6 - Preservation

Strategic Objectives

During the next 18 months (January 2005 – June 2006) the Preservation Work Package will pursue the following goals, which will contribute to the strategic objectives it has set itself for the project as a whole:

- establish a digital preservation testbed through implementing and integrating software tools;
- make progress on the automation (or at least semi-automation) of the processes of ingesting material into preservation environments through establishing approaches to improving the (semi-)automatic documentation of datasets;
- complete work on defining to enable the definition of attributes and functionalities that need to be represented;
- foster approaches to integrate preservation analysis and design approaches into system development methodologies;
- promote the adoption of preservation technologies in digital library development designs;
- raise the profile of digital preservation issues within the Digital Library Community; and
- increase our collaboration with other international researchers conducting research within the digital libraries and preservation communities.

The first 12 months of work undertaken by the Preservation Cluster built a foundation of activity, construction an environment of trust and collaboration between the partners in the cluster, and established mechanisms to enable us to integrate our preservation work into that of the cluster as a whole. The cluster is working towards developing threads of activity and method, which are ready for braiding together to form a preservation tool for DL activities. As our year 1 activity report makes evident, during the first twelve months the cluster worked to develop a foundation of work that would enable it to create a baseline for research into digital preservation in the context of digital libraries. We

made progress in developing an understanding of testbeds and the metrics and tools that we would need to put in place to assess preservation strategies, in understanding existing preservation storage environments and how these might or might not play a role in DLs, in better understanding the role of file formats and their typology and the relationship they have to preservation strategies. Our efforts to examine the definition of attributes and functionalities that need to be represented and to ensure that system development methodologies reflect preservation analysis and design issues have shown promise and these two activities will deliver results during the first six months of JPA2 and build the foundation for our core two new activities.

Tasks 4 and 5 were started under JPA1 and planned for completion towards the end of its programme of activity. We have continued working on these activities and they will form the foundation of JPA2. The first will complete at the end of the 6th month of the programme and the second at the end of the 12th month. Building on the work that we have so far completed the cluster intends to implement the testbed that it designed under JPA1 Task 1 and refined with work completed under Tasks 2, 3, and 4. In Task 7 below we intend to move forward with improving the process by which digital preservation systems can provide preservation standard documentation.

WorkPackage Activities and Expected Results

In the following we describe the activities to be carried out by the WP in the next 18 months and the expected achievement.

Task 6.4 - Documentation of Functionality and Behaviour Metrics.

This task has been working to define framework for documenting behaviour and functionality and to establish an overview of the attributes of functionality and behaviour that need to be represented and mechanisms for representing them. So far a workflow for the elicitation of these requirements was designed, resulting in a tree of objectives. These in turn, are assigned importance factors. Following the application of preservation actions, the results of these can be evaluated, allowing the comparison of different preservation strategies. Currently, the model is being conceptually integrated with the Testbed forum created as WP6-T1. Furthermore, information on the file format characteristics, Classification and Typology, which is currently being elaborated in WP6-T3 is evaluated and used to adapt the model. This work feeds directly into our plans for further work under JPA2 to develop and test preservation technologies within the DL environment.

Task 6.5 - Enabling the Integration of Digital Preservation Architectures

In this task the cluster has been working to establish the requirements for a preservation functionality modeling tool and to develop a method for modeling preservation functionality that could be integrated with more traditional system design and development methodologies. From a design point of view digital libraries are simply complex information systems, where the variance within the subset of digital libraries is so large, that it would be difficult, possibly even counterproductive, to use for their design a pronouncedly different design methodology than for other complex systems. As a result we have directed our attention to the question, how preservation aspects can be integrated into design methodology and its tools in general. While a number of possible strategies have been investigated, we have agreed to pursue a UML based solution. Our efforts during the first six months of JPA2 will be directed, to establishing positions within the existing UML language constructs to which "preservation extensions", based primarily upon the state diagrams of UML, can be attached.

Task 6.6 - Digital Preservation Testbed and Evaluation Framework

The overall goal of this Task builds on Tasks T1, T3 and T4 of WP6 of the previous JPA. It aims at implementing the testbed and evaluation metric design work performed in the past period, to allow automating major parts of the process of deciding about and evaluating preservation solutions is to further integrate, automate, and evaluate a framework for digital entity preservation by integrating and combining the testbed framework and the evaluation metrics developed during the first JPA of DELOS, with a specific focus on evaluation of the resulting framework in a set of real-world case studies at preservation institutions. This requires the development of specific tools to automate selected steps of the preservation process, such as ingest validation, preservation experiment set-up and control, preservation criteria definition, and verification of formal transformation, to support semi-automatic alternative evaluation. A second objective is to further investigate the evaluation metric, determining

the effect of small changes in the weight of the criteria on the final ranking of the alternatives, measuring, how well an alternative fulfils all criteria in comparison with an optimal preservation solution and which facets of an alternative to change to receive the highest impact. The goal in this context is to provide support for an automated sensitivity analysis of the resulting decision framework. Finally a third research objective is to evaluate our results in practice. Therefore, several real world preservation scenarios will be evaluated at partner institutions from different domains, specifically the OEAW-Phonogramm Archive, as well as OENB - National Library of Austria, and SUB - State and University Library of Goettingen (the latter two currently not being DELOS partners).

Task 6.7 - Digital Preservation Automated Ingest and Appraisal Metadata

The goal of this Task is to experiment with the integration and automation of appraisal and ingest of digital materials in sustainable digital repositories. The Task builds on the first 12 months work of the Work-package directly through using the outcomes of JPA1-WP6-Task1 (testbed), JPA1-WP6-task 2 (repository models), JPA1-WP6-task 3 (typology of file formats) and JPA1-WP6-task 5 (architecture modeling) as the foundation for further research into automating the processes of selection and ingest of materials into a preservation repository. JPA1-WP6 conducted (and continues to conduct) research into preservation testbed architectures (T1), reviewed current repository models (T2), examined file formats (T3), preservation strategies (T4), and architecture (T5) to construct a foundation for investigating mechanism for automating preservation processes. We seek now to combine the background work conducted under JPA1 with the results of ERPANET's Packaged Object Ingest Project, which examined the processes of ingesting digital objects (mainly documents) into a repository.¹ Even where it proved possible to use auto-extraction tools to generate the technical metadata (which for many formats was not possible because the tools were not available) the ingest process remained labour intensive. One core reason was the effort necessary to extract descriptive, structural and semantic information. This needed to be manually derived. The Task investigates the viability of automating several preservation processes as a mechanism for improving these processes and demonstrating the viability of automation more generally in the area of digital preservation and curation.

The research will focus on constructing a generic tool for extraction of semantically encoded metadata from any file type through the definition and prototyping of a tool for two or three file types and integrating this with tools for extracting technical metadata about file formats. In both instances the models will be based upon analysis of the metadata that the automation experiments demonstrate to be necessary in repository designs. From this tool it will be possible to establish a protocol or framework for creating semantic and structural metadata extraction tools. The framework should enable the 'factory-like' production of tools for other documentary file formats. These tools might be both automatic and semi-automatic (e.g. support interactive as well as batch application). The work will use LATEX, PDF, and XLS file formats for the case study and as the basis of generic frameworks.

Integration Activities

Integration is at the heart of the work of the preservation cluster—it is happening into three core areas. First, at a community level the work package has created links to the key research teams (e.g. National Archives in the Netherlands, Digital Curation Centre in the UK, InterPARES, National Library of New Zealand). Second, at DELOS level we are gradually increasing our cooperation with WP1, WP2, WP5 and WP7 in the areas of architectures, personalisation, knowledge extraction, and evaluation. Third, within the workpackage itself we have moved from having a series of poorly integrated tasks towards working as a cluster—this has been a gradual process

WP7 - Evaluation

Strategic Objectives

During the next 18 months (January 2005 – June 2006) the Evaluation Workpackage will pursue the following strategic goals:

¹ <http://www.erpanet.org>; final report on POIP available in July 2004.

- enabling discussions between evaluation specialists and developers and spreading evaluation knowledge among the latter,
- developing new evaluation models, methods and testbeds
- creating new evaluation toolkits for implementing these methods
- continuing existing evaluation initiatives, with a special focus on those aspects which are relevant to the digital library application area.

To achieve these goals, the following tasks are planned:

- Evaluation Forum
- INEX: An Evaluation Campaign for XML Information Access in Digital Libraries
- CLEF: An Evaluation Framework for Multilingual Information Access in Digital Libraries
- Digital Library Testbed Framework.

These four tasks are continuations from the first JPA (the last task originally was called 'Evaluation testbeds', to be started in month 13 of the Network).

Workpackage Activities and Integration

The main integration activities of this WP will involve research and application communities external to DELOS. Both the CLEF and INEX evaluation activities are involving a large number of research groups not only from Europe but also from the America, Asia and Australia. In this way, they are contributing to spreading DELOS knowhow and best practice in experimental evaluation to the wider interested communities. The other two tasks will primarily aim at integration within DELOS, as described in the following.

Task 7.1: Evaluation Forum

This task aims at facilitating communication about DL evaluation issues, and spreading knowledge about DL evaluation, thus supporting integration both to other DELOS WPs as well as to other researchers outside of the NoE.. For the first goal, a communication forum has been set up recently, and will be maintained at least throughout the whole duration of DELOS. In order to reach the second goal, we will compile surveys on existing DL evaluation methodology. This survey will consist of comprehensive collections of existing evaluation approaches, testbeds and toolkits.

Task 7.3: INEX: An Evaluation Campaign for XML Information Access in Digital Libraries

The aim is to provide the infrastructure and a framework to investigate the performance of systems and services that aim at providing effective access to XML-based DLs. To achieve this objective, two sets of complementary activities are planned: INEX campaigns on one side and research into appropriate evaluation methodologies and test-suites for both current and untested aspects of XML information access on the other. The INEX document mining track will be carried out in cooperation with the PASCAL Network of Excellence. In addition, INEX is organizing a mid-year workshop on elementary retrieval methodology as part of IR Fest in Glasgow, in July.

Task 7.4: CLEF: An Evaluation Framework for Multilingual Information Access in Digital Libraries

The provision of multilingual information access functionality is a key issue for Digital Libraries (DLs)..Two sets of complementary activities are planned: CLEF campaigns on one side, and research into new evaluation methodologies for untested aspects of multilingual information retrieval systems on the other hand. CLEF will be holding a one-day workshop in September in collaboration with the MUSCLE Network of Excellence. The CLEF group working on the evaluation of multilingual multimedia retrieval systems (ImageCLEF) intends to organize a joint evaluation activity with MUSCLE in 2006.

Task 7.5: A Digital Library Testbed Framework

This task will provide a standard testbed framework for comparative evaluation of digital library systems. The first step will be the definition of an integrated digital library evaluation model. Based on this theoretical framework, the Task will develop a testbed framework for digital library evaluation, by extending the Daffodil system (Daffodil is a front-end system consisting of a rich collection of services

for accessing federations of digital libraries). The result of this work can be used for verification of digital library execution versus a specification of digital library dynamics, as well as for quantitative analysis of digital library logs. The overall goal of the Digital Library testbed framework will be accomplished with the integration of contributions from WP4. It will integrate different services and will be applied to different applications.

Expected Results

For the four tasks of WP7, the following results are expected during JPA2:

- A collection of existing evaluation approaches, consisting of a list of publications related to the evaluation of digital libraries, testbeds and toolkits for DL evaluation.
- The INEX 2005 testbed, resulting from the INEX 2005 evaluation campaign. In 2005, INEX will comprise testbeds for the following tasks: adhoc retrieval, relevance feedback, natural language queries, heterogeneous collections, interactive retrieval, document mining, and multimedia retrieval.
- The CLEF 2005 testbed, resulting from the CLEF 2005 evaluation campaign. In 2005, CLEF will comprise testbeds for the following tasks: mono-, bi- and multilingual document retrieval on news collections, mono- and cross-language information retrieval on structured scientific data, interactive cross-language information retrieval, multiple language question answering, cross-language retrieval in image collection, cross-language spoken document retrieval, multilingual web track, and cross-language geographical retrieval
- A digital library testbed framework, which will be based on the Evaluation Computer and the Interaction Evaluation models, and also comprise a new logging standard for DL interaction events. A scientific primary data repository for digital library interaction logs will be set up, and appropriate tools for statistical analysis will be provided.

WP8 - Dissemination and Spreading of Excellence

The Spreading of Excellence activities fall into different categories, depending on the main objective of a specific activity, its contents and the expected recipients.

Scientific Dissemination

Thematic Workshops. This series of workshops is intended to provide the opportunity to European researchers to present results of on-going research activities and to exchange opinions and experiences in an informal and friendly environment.

Brainstorming Workshops. This series of workshops is intended to bring together top researchers from all the parts of the world, with the objective of defining "future research direction" reports in the field of Digital Libraries and to provide input for the definition of future research programmes in the domain of Digital Libraries both for the EC and national research funding agencies.

European Conference on Digital Libraries (ECDL). This conference has become the major European event on the subject of Digital Libraries, and DELOS is committed to maintaining its high scientific level, through the involvement of the most active research teams in Europe and in the rest of the world.

Education and Training

Summer Schools. This series of Summer Schools is intended to provide high-level courses on the domain of Digital Libraries and its underlying technologies. 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.

Awareness Events. This series of events (courses, tutorials, demonstrations, etc) is organized in collaboration with 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.

National/Regional Technology Transfer Workshops. This series of events is organized by Network members in cooperation with national application communities and industry, addressing themes of mutual interest. Particular attention will be given to the involvement of SMEs in each country.

Research Exchange Program. The exchange of researchers between organizations working on joint projects is one of the most effective ways to achieve integration of the working teams and exchange of skills and results. For "exchange of researcher" is meant the visit of a researcher from one DELOS member to another, for a period ranging from a few weeks to a few months. DELOS is providing some funding to cover the additional expenses that are required for those exchanges.

Visibility

Web site. The Network site is providing an overview of the main activities and events of DELOS, and also news and links about activities related to Digital Libraries, going on worldwide. The detailed descriptions about the activities of the Work Packages are provided by the web sites of the work packages themselves.

Electronic Newsletters. The Network is publishing a quarterly electronic newsletter, providing information and reports on DELOS events and on related activities of importance to the digital library community.

DELOS Awards. DELOS has established an award (in 2004 in the amount of 4 KEUR), named "DELOS Research Exchange Award", to be given each year to a young researcher who has authored a paper presented at the ECDL Conference. The award will recognise the achievement of the young researcher by offering the author a chance to spend a period of time (at least one month) at a DELOS European partner research organisation. In addition, DELOS is planning to give an award to the "Best ECDL Paper", based on the recommendations of the Program Committee. There are no conditions on this award, except scientific excellence. This award might be sponsored also by IEEE, as a way of cooperation between JCDL and ECDL.

WP9 - Assessment

This Workpackage has two specific objectives:

- to review and assess the overall activity of the Network and to provide recommendations
- to monitor and facilitate the integration process of the Network

The first objective will continue to be achieved by the Advisory Board of external experts, which is responsible for a periodic assessment of the activities of the Network, as described in detail in Section 8 of the Technical Annex.

The second objective will be achieved by guaranteeing the successful conduction of the second Joint Programme of Activities. As, already said, the second JPA is composed of a set of tasks; most of them to be carried out jointly by members belonging to different work-packages. Therefore, the successful conduction of a task will mean the satisfactory integration of the research activities of a number of members. It will be the responsibility of the Scientific Director and the Scientific Board to oversee the work progress and to guarantee the successful completion of the JPA.

The liaison persons nominated by each work-package will play an essential role for an effective coordination between the work-packages and will closely cooperate with the Work-package leaders, reporting on the status of integration, identifying critical aspects and possibly suggesting solutions concerning the:

- Definition and development of tools and solutions to enhance integration and information exchange between work-packages
- Definition and activation of discussion forums and events for information exchange between work-packages
- Definition and implementation of standards and interchange formats of the materials developed by the different work-packages

WP10 - Scientific Coordination

As in the first period, during the second period the scientific coordination workpackage will direct and supervise the scientific work of the Network. In particular, it will (i) organize the Network as a whole, (ii) supervise the scientific progress of the Network, (iii) ensure that all deliverables are available on time, (iv) create and maintain the conditions necessary for successful collaboration, (v) represent the Network in concertation with other scientific events.

The coordination will rely on the following management structures:

- Advisor Board (AB)
- Scientific Board (SB)
- Workpackage Steering Committees (WPSCs)
- Members General Assembly (MGA)
- Scientific Coordinator
- Administrative Coordinator

WP11 - Administrative and Financial Management

As in the first period, during the second period the administrative and financial management workpackage will ensure a strong and coherent administrative and financial management of the Network. The administrative and financial coordinator will also handle the reimbursement of the expenses incurred by the (invited) participants to events and meetings organized by WP8, WP9 and WP10 (e.g. workshops, national events, Advisory Board and Scientific Board meetings, etc.). For this reason, a large part of the money budgeted for those activities has been allocated to WP11 rather than to the WPs responsible for the events.

This Workpackage will, in particular, (i) handle all the administrative tasks connected with the NoE's activities, (ii) handle all the financial tasks connected with the NoE's activities, (iii) provide a global Intellectual Property Rights (IPR) frame for the whole participants, (iv) ensure institutional exchanges with the EC representatives, and (vi) promote the gender equality within the Network.