

**DIGITAL LIBRARIES:
Future Directions for
a European Research Programme**

**Brainstorming Report
San Cassiano, Alta Badia - Italy
June 13-15, 2001**

Table of Contents

1. Digital Libraries in the Future: A Grand Challenge Vision for the 6 th Framework Programme	4
1.1 Motivation	4
1.2 Grand Challenge.....	5
1.3 Follow-up actions.....	6
2. Digital Libraries in the Future: Why	7
2.1 Essential Technologies for the Development of Challenging Transformations in European Society.....	7
2.2 Uniqueness of the DL Environment.....	7
3. Digital Libraries in the Future: What	10
3.1 Research Hierarchy	10
3.2 A Conceptual Framework	11
3.3 Contents: 10-Year Goal and Technical Problem Areas	11
3.3.1 Collection Building	12
3.3.2 Collection Access and Navigation	12
3.3.3 Non-Traditional Kinds of Objects.....	13
3.3.4 Multilingual, Multicultural Collections.....	13
3.3.5 Collection Preservation	13
3.4 Management: 10-Year Goal and Technical Problem Areas.....	14
3.4.1 Basic System Architecture	14
3.4.2 Openness	15
3.4.3 Interoperability and Metadata	15
3.4.4 Scalability.....	15
3.4.5 Availability.....	16
3.4.6 Session-flow and Work-flow Management.....	16
3.4.7 Security.....	16
3.4.8 Quality.....	17
3.4.9 Digital Library Administration.....	17
3.5 Usage: 10-Year Goal and Technical Problem Areas	18
3.5.1 User Interfaces.....	18
3.5.2 Information Visualization	19
3.5.3 Personalization-Customization	19
3.5.4 Community Information Spaces	20
3.5.5 Multilingual and Multicultural Interaction.....	20
3.5.6 Collaboration.....	20
3.5.7 Universal Access	20
3.5.8 Multi-Channel Access	21
3.6 Applications and Impact	21
3.6.1 Application Areas.....	21
3.6.2 Socio-Economic Impact	22
3.6.3 Meta-Issues.....	22
4. Digital Libraries in the Future: How	24
5. Digital Libraries in the Future: Who	25
Appendix A – List of participants	26

Appendix B – Joint EC/NSF working groups..... 27

1. Digital Libraries in the Future: A Grand Challenge Vision for the 6th Framework Programme

1.1 Motivation

The European countries in their multi-millenary life have accumulated an enormous quantity of information, knowledge, experience, art treasures, etc. One only has to think of the art treasures contained in our libraries, archives and museums, or of the huge and precious collections of observational data in the areas of world exploration, sky observation, earth sciences, the environment, medicine, etc. accumulated during the past centuries. A huge amount of material also has been produced by the entertainment industry (TV, movies, music). A large part of these collections is currently available only on paper or in analogue form. This fact poses severe limits on their accessibility. In addition, other impediments (technological, physical, linguistic, cultural, legal, economic) have so far prevented citizens from taking full advantage of these existing valuable collections.

Nevertheless, recent advances in digital storage and digitization technologies are making the digital archiving of large collections both feasible and cost effective. Moreover, according to a recent report by Peter Lyman and Hal Varian at the University of California at Berkeley, the world currently produces between one and two exabytes (a billion billion 8-bit bytes) of information each year. Most of this information is in the form of images, sound, and numeric data; printed documents account for only 0.003% of the total. An increasing proportion of the information being produced is created, stored, and can be retrieved in digital form; more than 90% of this enormous annual output is now stored digitally. Yet, little of this information is made available through Digital Library collections.

Offering seamless universal and equitable access to the aforementioned collections will have a formidable impact on almost all citizens' activities (education, work, entertainment, culture, social activities, etc.). There is no doubt that by reducing barriers of distance, supporting timely sharing of resources and content delivery will greatly improve citizens' work productivity and quality of life.

Digital libraries represent a new infrastructure and environment that has been created by the integration and use of computing, communications, and digital content on a global scale. They are destined to become an essential part of the information infrastructure in the 21st century. They will make Europe's cultural and scientific heritage available to all European citizens, and sustain and preserve a universal collection of knowledge and creativity for future generations. New DL research, technologies and applications will greatly contribute to the increased use of distributed and networked information of all kinds and forms in Europe and the world.

1.2 Grand Challenge

In June 2001 the DELOS Network of Excellence organized a brainstorming workshop on "Digital Library Research Directions" (see Appendix A for the list of participants) with the objective of outlining the main research directions of the future European research programme in the field of Digital Libraries (DLs). Currently, the 6th FP (2002-2006) is being defined. The DELOS community believes that the active involvement of the European research community in its definition is of paramount importance. DELOS invited a number of prominent researchers to this meeting both from the Digital Library field as well from some important enabling technologies. The challenge was to outline advances in the enabling technologies that could have an impact on Digital Libraries and identify how such advances could contribute to the implementation of a new vision of Digital Libraries. The present report summarizes the results of three days of very intensive discussions.

After a fruitful discussion, the participants reached agreement on the following vision:

Digital libraries should enable any citizen to access all human knowledge any time and anywhere, in a friendly, multi-modal, efficient, and effective way, by overcoming barriers of distance, language, and culture and by using multiple Internet-connected devices.

After having reached agreement on what a Digital Library should be, the discussion concluded by specifying a grand challenge to be addressed within the 6th Framework Programme. The goal was to clarify through this the technical and social challenges associated with DL research. Even if the challenge's goals were not entirely reached in the end, they would provide high-level, technical benchmarks to help measure progress. The grand challenge envisaged is the following:

Establishment of an Initiative for an Integrated European Cultural Digital Library, which leads to the development of a comprehensive Digital Library of European history and cultural heritage.

Such an initiative can mobilize and motivate large numbers of people to work towards a specific goal having a strong positive impact on society, while at the same time significantly advancing scientific, technical, and humanistic activities. This initiative

- will help millions of citizens/students/learners to better understand the history and rich cultural heritage of each of the nations of Europe;
- will significantly advance many fields of the Digital Library research agenda;
- will provide a foundation of experience upon which other similar projects could be undertaken, especially large efforts with Europe-wide benefits;
- should operate in a distributed fashion with interoperability across collections and services as a requirement;
- should have built-in preservation;
- should have a clear plan for sustainability;

- should significantly improve synergy between national and EU-funded initiatives in the field of Digital Libraries.

1.3 Follow-up actions

In order to become a fully detailed research agenda, many of the research items outlined in this report require a much deeper understanding and investigation than is possible in a two-day meeting. To this end, it has been decided to establish a number of working groups jointly supported by the DELOS Network of Excellence and the National Science Foundation (NSF). The objective of each working group is to define a research agenda on a specific topic and to identify areas and activities for proposals to be submitted for funding under the future 6th Framework Programme (FP6) and the DLI2 programme of NSF, possibly establishing cooperation between EU and US researchers.

The ideal working group, which will be co-chaired by a preeminent EU researcher and a preeminent US researcher, will consist of approximately 10 members (5 from the EU and 5 from the US) and will meet 2-3 times alternately in Europe and the US. DELOS will fund EU researchers participating in the working group activities, and the NSF will do likewise with US researchers. The working groups are expected to start their activities by the beginning of 2002, and the final deliverables (which will be available before the end of 2002) are expected to be a white paper containing suggestions for future research directions in a specific topic, and proposals for future joint research activities.

The following topics have been selected for further investigation:

1. Spoken Word Digital Audio Collections
2. Information Extraction from Digital Libraries
3. Personalization and Recommender Systems in Digital Libraries
4. ePhilology: Emerging Language Technologies and Rediscovery of Past
5. Digital Imaging for Significant Cultural and Historical Materials
6. Preservation and Archiving
7. Test Collections and Performance Evaluation Methodologies
8. Actors in Digital Libraries.

Please see Appendix A for a list of the appointed co-chairs of each group.

2. Digital Libraries in the Future: Why

2.1 Essential Technologies for the Development of Challenging Transformations in European Society

The US President's Information Technology Advisory Committee (PITAC), in a report issued in 1999, identified several “National Challenge Transformations” as the essential prerequisites for enabling all citizens within their society to participate and fully benefit from the Information Age. In particular, transformation was considered necessary in the following areas:

- The Way We Communicate
- The Way We Deal With Information
- The Way We Learn
- The Way We Design and Build Things
- The Way We Conduct Research
- The Way We Understand the Environment
- The Way We Work
- The Way We Practice Health Care
- The Way We Engage in Commerce
- The Way We Offer Government Services and Information.

PITAC recognized the central role played by Digital Libraries in bringing about transformation in these areas, as they all assume or require Digital Library capabilities.

All participants at the meeting believe that the aforementioned transformations are crucial challenges for the European Union too, and their achievement depends significantly on the advancement of Digital Library technologies and capabilities.

2.2 Uniqueness of the DL Environment

One of the major issues that always arises in any discussion of Digital Library research is to define exactly what a Digital Library is and how it is different from other systems that it may erroneously be equalized with, e.g., a distributed, multimedia information system. Clarifying this distinction is important in order to identify what research needs to be done for the development of effective Digital Libraries that will not be done (or will be underfunded) in other governmental or commercial initiatives. The participants of the meeting believe that there are unique characteristics emphasized in Digital Library applications that lead to unique research agendas.

For the purposes of this document, the term “Digital Library” is used to capture everything that typically falls under a variety of terms, including “Digital Library”,

“Digital Museum”, “Digital Archive”, and others. There are three key characteristics that make a system a Digital Library and distinguish it from other kinds of systems:

- *Functionality*: It offers integrated services to a comprehensive digital collection of cultural or scientific information that is available primarily for reading and secondarily for expanding upon as well as annotation. Some of the features that are particularly emphasized in Digital Library information systems are
 - Rich information needs
 - Multiple sources of related information
 - Heterogeneous information
 - Rich data sources
 - Multimedia information
 - Defined user populations
 - Motivated users
 - Task-orientation
 - Domain-orientation
 - Cross-lingual access
 - Collaboration.
- *Purpose*: It is mainly used for learning and research.
- *Lifetime*: It provides access to information whose value is preserved across long periods of time.

Focusing specifically on the data and the users of information systems, an “information space” can be identified, with one dimension representing the level in which users and tasks are predefined and known in advance, and the other dimension representing the level in which the data has (known) structure. Given this information space, Digital Library applications can be distinguished from typical Web and database applications as shown in Figure 1:

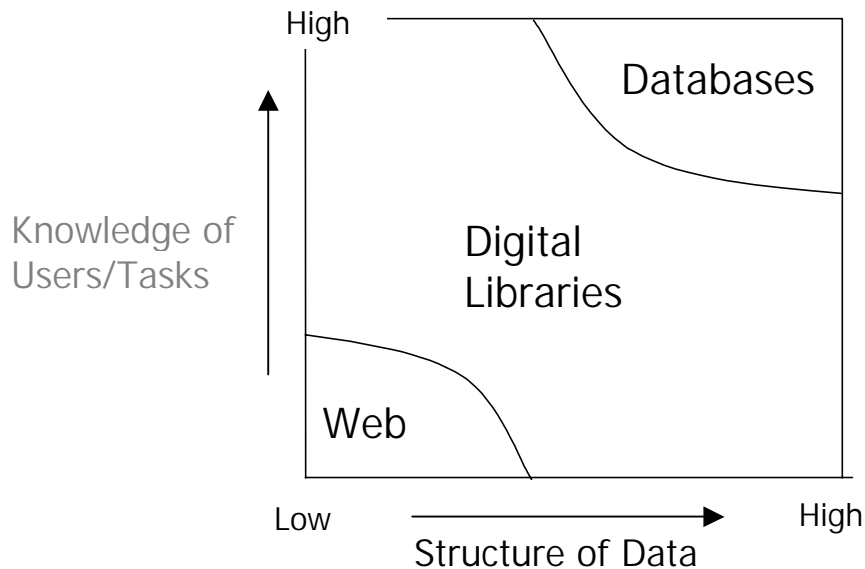


Figure 1: The Information Space for Digital Libraries

Although they receive much attention in the commercial world, typical Web search engines assume very little about users, tasks, and the data they deal with. Consequently, they occupy a relatively small part of the space. On the other hand, database applications (and some B2B Web applications) assume a great deal about users, tasks, and data. For example, the interaction with these systems is often limited to a few transaction types and data is typically defined using relational schemas. Hence, these applications occupy a small part of the space as well. The rest of the space can be viewed as belonging to Digital Library applications. In this part of the space (which is by far the largest), information systems attempt to exploit knowledge about the users, tasks, and domain to improve access, but retain the flexibility of ad-hoc querying, filtering, presentation, etc. that is characteristic of many Web-based applications. This mixture of characteristics leads to many unique research challenges and interesting test-bed applications.

Another interesting comparison is that between Digital Libraries and the GRID. The latter is conceived as tying together heterogeneous computation and data resources through the use of middleware, and then applying techniques such as data mining and others on these resources to infer higher-level knowledge. This is essentially part of the functionality offered by a Digital Library system; hence, the GRID can be considered as a special case of a Digital Library.

3. Digital Libraries in the Future: What

3.1 Research Hierarchy

To obtain a more organized vision of the research that we¹ view as critical for the future, we have used the following ‘Research Hierarchy’ as a template throughout this document.

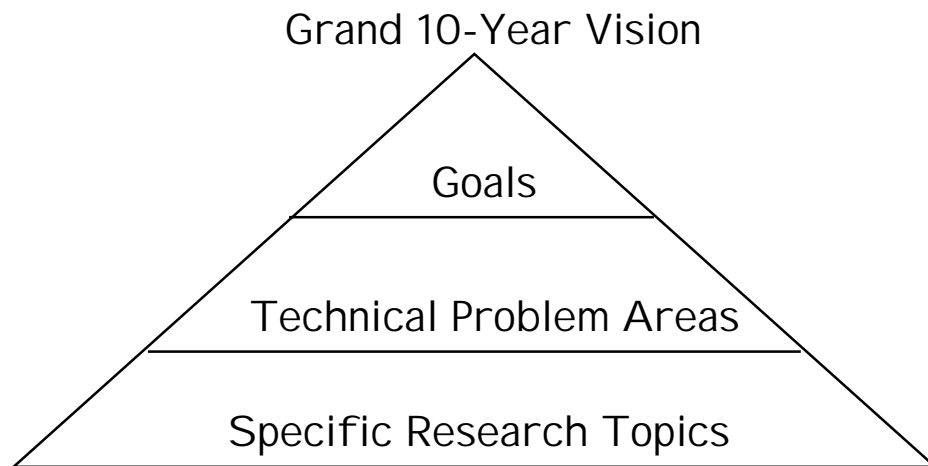


Figure 2. Research Hierarchy

At the top of the hierarchy, there is a ‘Grand 10-Year Vision’ for the entire area of Digital Libraries. Achieving this vision requires major advances in several aspects of Digital Library systems but also implies significant changes in the way we search for information, for all levels of research and learning. At the next layer, there is a small number of ‘Goals’, one for each of the major components of a Digital Library system or its environment. These can be thought of as more specialized ‘10-Year Visions’.

At the third layer and under each ‘Goal’, there are several ‘Technical Problem Areas’ that require attention, where major progress is needed to achieve the corresponding ‘Goal’. Finally, at the leaf level and under each ‘Technical Problem Area’, there are ‘Specific Research Topics’ within the parent ‘Technical Problem Area’ on which novel research work is necessary.

The shared 10-year Grand Vision for Digital Libraries, which has been described in Section 1, can be re-stated here in different words:

Anyone should be able to receive all information and services they want from any Digital Library, anytime and anywhere, in the most efficient and effective way.

¹ For brevity in this section, “we” is used instead of “the participants of the meeting”.

3.2 A Conceptual Framework

In order to become more specific, a general conceptual framework for Digital Library systems was defined, as depicted in Fig. 3:

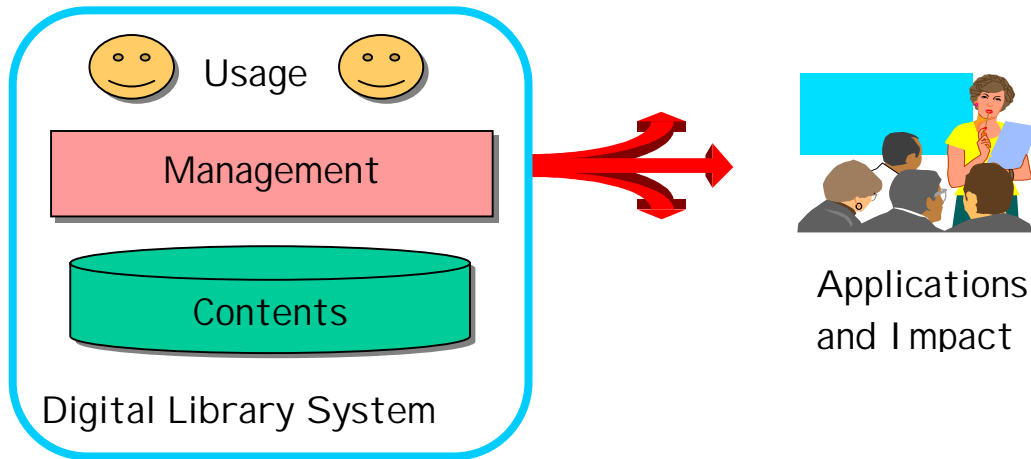


Figure 3. A Conceptual Framework for Digital Libraries

On the left-hand side, there are the three major components of a Digital Library system. At the bottom are the contents of the Digital Library. On top of it is the core system, responsible for the management of the contents and for providing the necessary functionality. At the front-end is the user interaction component, dealing with all aspects of the interface between the users and the system.

For each of these three components of a Digital Library system, we establish a ‘10-Year Goal’ below, and then analyze the research work that we see necessary to reach it. Clearly, any new approach, solution, or enhancement in each of these components affects some or all of the others as well, generating more related research problems. Our analysis places within each dimension the research problems whose primary motivation lies there.

On the right-hand side of the above figure, there is the outside world, the general society. This represents all applications that could benefit from advanced Digital Library systems and the precise impact the latter would have on the former. We analyze this dimension as well and identify some key directions of work that should be followed in the future.

3.3 Contents: 10-Year Goal and Technical Problem Areas

Starting from the bottom of a Digital Library system (its contents), the following expresses what we see as the relevant high-level vision for the next ten years:

Creating high-quality, semantically rich, comprehensive information collections, usable for long periods of time.

To achieve this goal, we identify four major technical areas where several problems remain unsolved and require attention:

- Building an information collection
- Accessing an information collection and navigating through it
- Dealing with non-traditional kinds of objects in a collection
- Dealing with multilingual, multicultural collections
- Preserving an information collection.

In separate subsections below, we outline the particular research issues that we see as most critical in each of these areas.

3.3.1 Collection Building

Although building the information collection appears to be a rather mundane task, it is a critical process and anything that can be done to facilitate it is important. The key research topics are the following:

- Information acquisition: Automatically acquiring the primary contents of the Digital Library.
- Information analysis and extraction: Generating “meta-information” on top of the primary contents. Example processes for generating such information include annotation, link creation, summarization, classification, and others. To a large extent, the resulting information tends to have value comparable to the primary/raw data.
- Situated information organization: Organizing both the primary and the secondary information in ways that are appropriate for specific situations, e.g., specific types of usage, specific conceptual approaches appropriate for different user groups, etc.

3.3.2 Collection Access and Navigation

Searching through the data in a collection is the centerpiece of all required processing, so it is affected significantly by all novelties envisioned in future DL systems. The main challenges follow:

- Efficient search algorithms and structures: With so many new forms of data and their combinations, new search algorithms and structures need to be developed that can take advantage of the particularities of the data, access it appropriately, and provide results efficiently. The difficulty of this task is further exacerbated by the expanded nature of the searches users would like to perform on the data.
- Search optimization: The complexity of data search and manipulation in DL systems demands new approaches to query optimization. Especially critical is the issue of size and cost estimation, which is an area with no prior work for most forms of DL content. Advances in this area will also help in providing sophisticated pre-execution user notifications with respect to cost and result size.

3.3.3 Non-Traditional Kinds of Objects

Digital Libraries of the future will need to deal with several more kinds and forms of information than currently. Of critical importance are the following:

- **Scientific data collections:** In addition to textual information, which has been the primary focus of Digital Libraries until now, raw scientific data collections should be emphasized as well, for a more direct impact on scientific experimentation.
- **Simulation models:** Not only scientific data, but the scientific processes themselves should become part of Digital Libraries. In particular, simulation models should be stored in Digital Libraries and become available through them, either as a commodity or as a service. Scientists should be able to compose these in meaning scientific workflows, feed them with appropriate data, and run the corresponding experiments, all as part of interacting with a Digital Library. Thus, the entire spectrum of scientific discovery, from initial conception of ideas, to experimental exploration, to publication of the final results will be served through Digital Libraries.
- **Combinations of text, video, audio, images, structured data, and other forms:** Digital Libraries should become able to manage all available forms of information in an integrated fashion to support the needs of their users. So far, much effort has been put into building mono-media Digital Libraries (text, video and audio). In the near future significant effort should be devoted to building truly multi-media Digital Libraries as very few of the on-going projects deal with this issue.

3.3.4 Multilingual, Multicultural Collections

A specific kind of non-traditional object collection that requires particular attention due to its special importance in the world of Digital Libraries is that of multilingual and/or multicultural collections. In the current era, the global nature of science and culture is more apparent than ever, so Digital Libraries need to become able to support studies of this nature. The basic issues in this direction appear to be two:

- **Culturally-driven information translations:** Information should be available in many languages and within the framework of many cultures. As storing it in all required forms is not possible, techniques for translating between languages while taking into account cultural backgrounds should be developed.
- **Information and meta-information:** To achieve the above, particular linguistic and cultural (meta-) information should be identified that needs to be stored and used during the translation process and thereafter.

3.3.5 Collection Preservation

An important area that is only now becoming part of research agendas is that of preservation of Digital Library collections, which is intimately related to the “valuable at depth of time” aspect of a Digital Library definition. Two main technical challenges are identified:

- Software and information migration: As technology moves forward, techniques should be developed to (semi-) automatically migrate the contents and processes of a Digital Library to new environments so that they remain available to their users. For software, this may be migration to new hardware platforms or new programming languages, for example, while for information, this may be migration to new data formats or semantic conceptualizations.
- Translation algorithms and techniques: Information translation is an important and particularly hard component of migration, so special attention should be paid to the development of generic translation algorithms and high-level translation specification languages.

3.4 Management: 10-Year Goal and Technical Problem Areas

Moving on to the kernel component of a Digital Library system, the one related to the system's 'Management', the following expresses what we see as the relevant high-level vision for the next ten years:

Developing self-sustainable and expandable DL systems, offering high-quality information and services.

To achieve this goal, we identify several major technical areas where several problems related to the system's architecture remain unsolved and require attention:

- Basic system architecture
- Openness
- Interoperability and metadata
- Scalability
- Availability
- Session-flow and work-flow management
- Security
- Quality
- DL administration.

In separate subsections below, we outline the particular research issues that we see as most critical in each of these areas.

3.4.1 Basic System Architecture

The current typical client-server and 3-tier architectures are not adequate to provide the functionality implied by the advances expected in the remaining architectural issues. Specific effort is needed in exploring novel architectures, particularly these two kinds:

- Component-based architectures
- Multi-tier architectures.

3.4.2 Openness

The Digital Libraries of the future will be ever-expanding systems. An open architecture implies that the overall functionality of the Digital Library will be partitioned into a set of well-defined services. A Digital Library will consist of smaller independent systems that will each provide different functionality or access to different contents. Hence, work is needed in the following areas:

- **Plug-and-Play flexibility/modularity:** When a new service is added to the system functionality, a new component should be able to come up and work. That is, it should become possible for individual systems to be easily plugged into a Digital Library system as components.
- **Auto-description, auto-registration, auto-configuration:** An important aspect of providing the required openness is the ability for systems (and information collections) to be self-describing so that, when plugged into a system, they can be (semi-) automatically registered and configured. Any other manual process will not scale to the level required.

3.4.3 Interoperability and Metadata

Given the non-monolithic nature of future Digital Libraries, interoperability is at the core of systems requirements. There are several research issues that arise in this context, but the most critical one appears to be the following:

- **Metadata correlation:** Metadata of information and software interfaces should be (semi-) automatically correlated so that syntactic and semantic heterogeneity can be addressed. This will allow for software to interact with other software and information to be moved from one form to another within an open, multi-component environment.

In addition to research work per se, some infrastructure should be built to facilitate both the development and operation of interoperable systems. This should take two forms primarily:

- **Registries:** Registries of meta-data, meta-services, and meta-mappings (i.e., generic mappings between data formats and schemas) should be established.
- **Conversion tools:** Software for data conversion should be developed and made available to the community as a resource shared and used by everyone.

3.4.4 Scalability

Scalability will be an important aspect of future Digital Libraries systems, given their ever-growing nature. Moreover, scalability will have to be exhibited at the levels of users, system components, and contents. To support scalability in the new environment, work is required in the following areas:

- **Decentralized architectures:** The most appropriate architectures should be identified that will support scalability. Particular effort should be put on the

investigation of Peer-to-Peer architectures, the GRID architecture, and cluster architectures, as well as on the conception of brand-new architectures.

- Performance prediction: An important issue around scalability is estimation of the performance impact that the addition of a new user or a new component will have on the system.

3.4.5 Availability

Another important characteristic of a Digital Library system is availability. Particular issues that arise in developing highly available systems, based on an open decentralized architecture, include the following:

- Dynamic reconfiguration: When some components of the system fail, automatic mechanisms should be triggered to compensate for the failure. Requests for information should be redirected so that the faulty parts are avoided.
- Replication: A major prerequisite for effective reconfiguration is replication, for which new methodologies should be developed, for a Digital Library environment. Extensions to mirroring methods will also help improve performance.

3.4.6 Session-flow and Work-flow Management

Accessing the information and services offered by a Digital Library may become quite involved. In that sense, managing the flow of the session or work of the user is critical. This is an area with almost no prior related work and requires effort in several directions:

- Modeling: The appropriate models for interacting with a Digital Library should be identified, so that all other aspects of session-flow management can be based on them.
- Correctness and consistency: Based on the models conceived, appropriate semantics for correct and consistent session/work-flows should be defined.
- Long and interoperable sessions: The above should be investigated particularly for long and interoperable sessions, which will be most common and most difficult to handle. Of significant importance will be the distinction between persistent work-flows, which are canned/prepared paths of interaction with the Digital Library, and ad hoc session-flows, which are arbitrary sequences of events that users will follow, constructing them dynamically.
- Extensibility: If Digital Library operations can be described, edited, and used to generate work-flow handling code, then changes to a Digital Library can be made easily, without requiring extensive programming. Any advances in this direction will have great impact.

3.4.7 Security

Security is another critical issue around Digital Libraries. Three of the typical aspects of security (privacy, anonymity, and authorization) appear to be addressable by standard

approaches, not affected by any particular characteristic of Digital Libraries. For other aspects, specialized approaches should be identified:

- Integrity: Due to the complexity and richness of a DL environment, enforcing and guaranteeing the integrity of its information contents requires attention.
- Confidentiality: The same attention should be placed on guaranteeing the confidentiality of users' actions.
- Digital rights specification languages: A very important aspect is how to protect the Intellectual Property rights of the owner of the digital material. Specific work is necessary in developing specification languages for expressing access rights (and possibly fees) for all forms of digital material.

3.4.8 Quality

The quality of services offered by a Digital Library is very critical to its viability. Yet there is currently no understanding of the concept of 'quality' for the services of a Digital Library. Several issues need to be studied in this direction:

- Quality criteria: These should be developed formally so that the meaning of quality may be identified. They should consist of specific criteria related to information correctness, information completeness, information age, guaranteed service termination, information and service cost, and possibly others.
- Metrics: Given criteria for quality, ways to measure them should be identified.
- Estimation: Given metrics for quality criteria, techniques to estimate/approximate them should be developed, as accurate measurements will be prohibitively expensive.
- Quality-based processing: Often a Digital Library must process requests based on quality criteria. These criteria will be imposed by the user or will be enforced by the system in general. Optimizing and executing requests based on such criteria is an extremely difficult but also necessary problem to solve.
- Quality-oriented metadata: For quality to be taken into account during optimization or processing within a Digital Library, the metadata should be enhanced with quality-related pieces of information.

3.4.9 Digital Library Administration

System administration is a rather mundane task, yet its semantics within a Digital Library environment is missing. A "Digital Library Administrator" controls both ends of the overall system, including the design, population, and organization of the contents of a Digital Library, as well as the definition of its individual users and user communities. The concept is very similar to that of the "Data Base Administrator". This area may not require fundamental research, but it certainly demands the development of administration tools for work to move forward. One key issue is to establish standards for logging activities in Digital Libraries, so that, for example, systems can be compared, usage can be analyzed, and performance can be understood.

3.5 Usage: 10-Year Goal and Technical Problem Areas

The highest-level component of a Digital Library system is related to the system's usage. For this area, the following expresses what we see as the relevant high-level vision for the next ten years:

Provide optimal user experience in Digital Library interactions, i.e., support users in accessing Digital Libraries and ensure that they obtain the desired information in the best possible way.

To achieve this goal, we identify several major technical areas where several problems remain unsolved and require attention:

- User interfaces
- Information visualization
- Community information spaces
- Multilingual and multicultural interactions
- Personalization and customization
- Collaboration
- Universal access
- Multi-channel access.

In separate subsections below, we outline the particular research issues that we see as most critical in each of these areas.

3.5.1 User Interfaces

Despite much work in the area of user interfaces that affects a great variety of applications, little attention has been paid to some specific characteristics of interacting with Digital Libraries that raise several new issues requiring solutions. The most important of those are the following:

- Integrated multi-paradigm access: Digital Libraries will manage information residing in a variety of data-centric systems ranging from relational databases to unstructured documents to non-textual, multimedia data. The established paradigms for interaction with any one of these systems differ drastically from those of the others. Hence, either new paradigms should be devised that subsume the existing ones, or techniques should be developed to support the integrated use of the existing ones. In that direction, there are three particular questions that appear critical. The first one is related to the syntax and semantics of user-level languages that are appropriate for posing multi-paradigm requests. The second one is related to the semantics of the correct answer when it is formed from a combination of answers from diverse systems. The second one is related to the efficient processing of multi-paradigm requests.
- Task-oriented access: Interaction with a Digital Library can take many forms depending on the task that is being performed. A universal, generic interface is

bound to be ineffective, so effort should be put into developing interfaces that facilitate particular tasks.

- User interfaces generation: Interface description languages (like UIML - user interface markup language) should be developed so that interface families are described, and then particular ones generated from those specifications, suitable for various combinations of devices. This can support work on generalizing understanding of interfaces, as well as personalization.

3.5.2 Information Visualization

Visualizing information in Digital Libraries presents several difficulties, particularly due to the variety of what can be visualized. The key specific question that needs to be addressed is developing techniques and systems that support visualizations that are dependent on the nature of the information visualized, both at the level of the actual contents and at the level of the meta-contents.

3.5.3 Personalization-Customization

Personalization and customization of interaction and overall user-experience with a Digital Library remains a critical issue, as Digital Libraries must match if not surpass regular libraries in these aspects. Work in this area needs to proceed in several directions:

- Explicit and implicit profiling: Personal profiles of interaction with a Digital Library can be identified either explicitly or implicitly. Examples include thorough initial interviews, or thorough observations of past user behavior and data mining on the findings, respectively. The effectiveness of each approach should be studied and the appropriate combination of them should be identified.
- Static and dynamic profiling: Both forms of profiling can be done either statically (e.g., at user registration time only) or dynamically (e.g., throughout the user session and throughout the operation of the system). The challenge is to develop techniques that will support the dynamic generation of profiles, so that changes in user behavior are reflected in the system reaction as well.
- Personal annotations: As with regular books and other printed documents, users should be able to generate personal annotations about the digital objects they are interacting with, which will appear with the same objects in subsequent uses by the same users. Space-efficient storage of these annotations and intelligent processing of requests that takes into account the existence of these annotations are challenges that require attention.
- Person-dependent system behavior: Digital Library systems should deliver both content and services to their users according to the profiles of the latter. Techniques for achieving that efficiently need to be developed. In addition to the personalization/customization of what is delivered to a user, and of the type of interaction with the Digital Library that is supported, equally important is the personalization/customization of the interpretation of the user requests, an area with almost no existing work.

3.5.4 Community Information Spaces

All the issues mentioned can be extended beyond the personal level, to the level of communities. Digital Libraries systems need to identify and create “Community Information Spaces”, where different users belonging to different communities can observe a different behavior of the system based on such ties. Work in this area needs to explore the following:

- Implicit and explicit community definition: This represents a difficult clustering problem that requires specific attention.
- Community annotations: In order to allow the members of a community to collaborate through the Digital Library resources, it is very important to extend the support of annotations to the community level.
- Ratings: Opinions of some in a community can help guide others, as in peer review and other scholarly publishing processes. Reconciliation of diverse opinions is a challenge that needs to be overcome to achieve “community ratings”.

3.5.5 Multilingual and Multicultural Interaction

Particularly important communities are those defined based on the native language and/or native culture of the users (i.e., the culture in which they were brought up). These communities are predetermined and the effect that they should have on system behavior is significant and much further reaching than that of other types of communities. Much work is necessary to support language-dependent and culture-dependent user requests as well as language-dependent and culture-dependent content and service delivery.

3.5.6 Collaboration

An important new aspect that is raised by Digital Libraries with respect to collaborative systems is “synchronous Digital Library visits”. A platform should be developed to permit multiple users to interact with a Digital Library simultaneously, each one being aware of the presence of the other and being able to interact with each other as well. This will approximate the experience of visiting a traditional library or museum, and the educational benefits that non-individual, collective visits may have.

3.5.7 Universal Access

Access to Digital Libraries should be universal. This can be interpreted as universality in three dimensions: people (access by everybody), location (access from everywhere), and devices (access via everything, e.g., regular computer screen, palm organizer, etc.). Work is needed to increase the level of inclusiveness of Digital Libraries in all directions.

3.5.8 Multi-Channel Access

Universality of access with respect to the device dimension is especially critical when temporal aspects are introduced and users are allowed to access a Digital Library using different devices at different times. The main challenges are as follows:

- Persistent sessions across multiple devices: Techniques should be developed to maintain user sessions persistently even when they move around among diverse devices. They should be able to pick up their work from where they left off.
- Device-dependent content and service delivery: Techniques should be developed to support delivery of both content and services to the users that is dependent on the device where they are going to be delivered. A request posed from one device should have a different response (in terms of visual abstraction level) when that is viewed from the same device or a different one with different capabilities.

3.6 Applications and Impact

Having completed the directions for future research work in the various layers of a Digital Library system, we move on to the overall environment where Digital Libraries operate and examine the ‘Applications and Impact’ of this technology. By the nature of the topic, there is no 10-year grand vision here, but there are several technical and non-technical areas that require attention:

- Application areas
- Socio-economic impact
- Meta-issues.

As before, in separate subsections below, we outline the particular issues that we see as most critical in each of these areas.

3.6.1 Application Areas

The technology of Digital Libraries will help many other areas of scientific, engineering, or business endeavor. Some of the application areas that could greatly benefit from the adoption of Digital Libraries technologies are:

- Education
- Medicine
- Entertainment
- Cultural Heritage
- Science & Technology
- Government
- Environmental.

Among them, one that appears to be most critical is the use of Digital Libraries in Education, as it affects essentially everyone. Particular emphasis should be put on the following issues:

- **Impact:** Cognitive studies should be conducted to quantify the impact of using Digital Libraries in Education and how it affects learning by various categories of users (e.g., high-school students, college/university students, or distant learners).
- **System needs:** Any particular demands placed upon Digital Libraries within an educational environment should be studied, their impact on systems aspects should be identified, and the required technical solutions should be devised.
- **Infrastructure:** Building some Digital Library services on a variety of educational topics and in several parts around the world is necessary to make Digital Libraries effective and widely accepted as an important medium in the educational process.

3.6.2 Socio-Economic Impact

One aspect of Digital Libraries that is often ignored from relevant studies is that of the socio-economic impact that they may have. We identify three important issues in this direction:

- **Business modeling:** For Digital Libraries to become common place, they have to operate based on meaningful business models just like their physical counterparts. This is even more critical here, as many Digital Libraries will provide content that is privately owned. Identification of such business models and studies of their effectiveness is a priority.
- **Sustainability:** Somewhat related to the above is the issue of sustainability. Digital Libraries must remain current. Hence, mechanisms should be identified to fund the continuous renewal of material in them and maintain users' awareness of their offerings.
- **Copyright:** Issues of copyright are notoriously difficult to solve, especially because they appear to be different from case to case. Some effort should be put on possibly identifying a small number of standard approaches that could be used in several cases.

3.6.3 Meta-Issues

Several technical and semi-technical issues around Digital Libraries exist that are not directly related with the internals of a Digital Library system itself. Three of these appear to be most critical:

- **Methodologies:** Currently there is no established way on how to develop a Digital Library. This refers not only to the software that is necessary, but also to the contents collection and/or acquisition, the daily management of the environment, dealing with change, and several other aspects. Development of such methodologies and understanding their effectiveness in different situations is a critical prerequisite for introducing Digital Libraries in all but the most advanced environments, e.g., remote areas or non-technical applications.
- **Standardization:** A critical tool towards Digital Library development, especially with respect to interoperability, is the development of standards. Despite the

existence of many of them for various aspects of content storage, software interfaces, or metadata conceptualization, many more are needed to capture the richness of the Digital Library environment.

- Digital Libraries as subsystems: Not only will a Digital Library system be component-based, but the entire system will often serve as a component of a much larger environment as well. Work should be done on finding ways to facilitate this, including aspects of Digital Library interfaces to external systems.

4. Digital Libraries in the Future: How

The workshop recommended the establishment of a large Initiative for an Integrated European Cultural Digital Library. Below are some thoughts on how to structure such an Initiative. Some of them build on the US NSF development of a National SMETE (Science, Mathematics, Engineering, and Technology Education) Digital Library (NSDL).

The Initiative should be structured as a cluster of projects organized by tracks:

- Core Integration Track
- Collection Track
 - By institution type (e.g., national archives)
 - By genre (e.g., news, documentaries, virtual tours)
 - By area (e.g., art, history)
- Services Track
- Research Track.

To pursue the full range of research questions embraced by Digital Library technologies, small projects working in conjunction with larger ones would be particularly effective. (Note that there are significant research components in the first three tracks as well.)

This type of organization should produce diversity without duplication, and coordination without stifling the Initiative. Both will be needed to pursue the different kinds of interoperability. Such a structure would also allow several projects to work around a particular test-bed, allowing the digital collection to coordinate the research community.

It was also suggested that, within the overall funding cycle, different projects should be organized with different periodicity, to provide successive waves of research, building on previous results.

An additional beneficial outcome of this Initiative should be the creation of a cooperative community of Digital Library researchers.

5. Digital Libraries in the Future: Who

The workshop stressed the importance of strategic partnership - private, public, national, European, and international - to give synergy to Digital Library research. In order to build such an important Initiative like the one proposed by this workshop, co-investment is required. In addition, coordination of national and EU-level efforts in the Digital Library field is also needed.

There are several reasons why Digital Library research should be primarily funded by and conducted at the European (Union) level, and sometimes at an even broader level, where Europe cooperates with other entities at its level:

- Information of interest to be stored and maintained within a Digital Library is by definition multinational, e.g., cultural and scientific heritage is essentially global.
- As a result of the above, Digital Libraries will be predominantly international establishments.
- The Digital Library field is very rich and complex from a technical standpoint, so expertise in all of its aspects is naturally international.

The core stakeholder groups are:

- Memory-based Organizations (Libraries, Archives, Museums)
- Universities and Research Organizations
- Broadcasting Industry
- Electronic Publishing Industry
- Software Industry
- Telecommunication Industry.

Partners will help in covering the cost of research in a variety of ways. Some will provide funding; others may bring their own expertise as an in-kind contribution. And yet others may bring existing collections, authentic users, and a valuable understanding of real-world problems to the table. There should be diverse ways of participating.

All meaningful types of collaborations and co-funding should be encouraged to meet the objectives of the Initiative. Some problems, mainly technological, suggest collaboration with the US; other problems, like solving multi-lingual and multi-cultural access challenges, by their nature, suggest certain types of collaboration (e.g., with initiatives in Asia, Japan, China). Some types of collections (e.g., space data, geographic information) suggest involvement of international agencies, such as the European Space Agency. In summary, co-funding to meet research objectives and increase leveraging of Community funding is highly desirable.

The DELOS NoE can play an important role in undertaking many of the appropriate actions in order to mobilize the core stakeholders, in defining and managing the contents of the Initiative, and in carrying out some of its research activities.

Appendix A – List of participants

Elisa Bertino (University of Milan, Italy)
Vittore Casarosa (IEI-CNR, Italy)
Gregory Crane (Tufts University, USA)
Bruce Croft (University of Massachusetts, USA)
Alberto Del Bimbo (University of Florence, Italy)
Dieter Fellner (Technical University Braunschweig, Germany)
Paul Fiander (BBC, UK)
Edward Fox (Virginia Tech, USA)
Norbert Fuhr (University of Dortmund, Germany)
Yannis Ioannidis (University of Athens, Greece)
Traugott Koch (Netlab, Denmark)
Ronald Larsen (University of Maryland, USA)
Patricia Manson (European Commission, Luxembourg)
Pier Giorgio Marchetti (European Space Agency, Italy)
John Mylopoulos (University of Trento, Italy)
Erich Neuhold (FhG-IPSI, Germany)
Paolo Paolini (Politecnico of Milan, Italy)
Mike Papazoglou (University of Tilburg, The Netherlands)
Laurent Romary (INRIA Lorraine-Loria, France)
Hans-J. Schek (ETH-Zurich, Switzerland)
Alan Smeaton (University of Dublin, Ireland)
Costantino Thanos (IEI-CNR, Italy)
Howard Wactlar (Carnegie Mellon University, USA)

Appendix B – Joint EC/NSF working groups

1. Spoken Word Digital Audio Collections
EU co-leader: Steve Renals (University of Sheffield)
US co-leader: Jerry Goldman (Northwestern University)
2. Information Extraction from Digital Libraries
EU co-leader: Yannis Ioannidis (University of Athens)
US co-leader: David Maier (Oregon Health and Science University)
3. Personalization and Recommender Systems in Digital Libraries
EU co-leader: Alan Smeaton (Dublin City University)
US co-leader: Jamie Callan (Carnegie Mellon University)
4. ePhilology: Emerging Language Technologies and Rediscovery of Past
EU co-leader: Susan Hockey (University College London)
US co-leader: Gregory Crane (Tufts University)
5. Digital Imaging for Significant Cultural and Historical Materials
EU co-leader: Alberto del Bimbo (Florence University)
US co-leader: Ching-chih Chen (Simmons College)
6. Preservation and Archiving
EU co-leader: Seamus Ross (University of Glasgow)
US co-leader: Margaret Hedstrom (University of Michigan)
7. Test Collections and Performance Evaluation Methodologies
EU co-leader: Norbert Fuhr (University of Dortmund)
US co-leader: Ron Larsen (University of Maryland)
8. Actors in Digital Libraries
EU co-leader: Jose Borbinha (National Library of Portugal)
US co-leader: John Kunze (University of California, San Francisco)