

Digital libraries at a crossroads

Yannis Ioannidis

University of Athens, Department of Informatics and Telecommunications, Athens, Greece
e-mail: yannis@di.uoa.gr

Published online: 22 July 2005 – © Springer-Verlag 2005

Abstract. *Digital libraries* constitute a relatively young scientific field whose life spans roughly the last 10 years. Despite its youth, the field has made a long journey from its initial conception to the present state of the art and has reached a level of maturity that warrants taking stock of its course so far and paving a path for the immediate and long-term future. In this paper, we analyse the recommendations of several activities that were initiated by the DELOS Network of Excellence under the Fifth Framework Programme of the European Commission and demonstrate that there are several issues that arise in similar form in most digital library efforts. We thus conclude that generic *digital library management systems* that address all these issues in a common fashion can and should be developed, so that future digital library efforts are able to build on such platforms and focus only on the hard issues that specialised functionality raises.

Keywords: Digital libraries – Digital library management systems – Digital library research roadmap

1 Introduction

Digital libraries represent the meeting point of a large number of disciplines and fields, i.e., data management, information retrieval, library sciences, document management, information systems, the Web, image processing, artificial intelligence, human-computer interaction, and others. It was only natural that the first 10 years of their existence were spent mostly on bridging some of the gaps between the disciplines (and the scientists serving them), understanding what ‘digital library functionality’ is supposed to be, and integrating solutions from each separate field into systems to support such functionality. These goals have been achieved through much

exploratory work, primarily in the context of focused efforts that have devised specialised approaches addressing particular aspects of digital library functionality. For example, the ARTISTE project from Europe’s Fifth Framework Programme focused on how to develop an integrated analysis and navigation environment for art images and analogous multimedia content, the COLLATE project from the same Programme focused on how to deal with old film libraries, while the Alexandria Project from NSF’s DLI-1 and DLI-2 Programs focused on geospatially referenced multimedia material. For the most part, every effort so far has been distinct and, in some sense, isolated from the rest. Each project has started from scratch to build a system supporting the particular needs specified in the project’s description. Nevertheless, looking back at the individual achievements of all the projects, one can see clearly that there is substantial commonality among many of them.

The field of digital libraries is at a level of maturity that did not exist 10 years ago. Substantial knowledge and experience have been accumulated.¹ Moreover, one may now claim that there is indeed a ‘critical mass’ of scientists coming from very diverse backgrounds that form a digital library research community. All the individual activities of the past seem to have come together bringing digital libraries to a crossroads. How does the field move forward from this point?

This paper attempts to offer a relatively detailed roadmap on how to proceed from the current state of the art to the next generation of digital libraries. It identifies some of the principal research problems that need to be

¹ Instrumental in the birth and growth of the field have been the funding opportunities generated by the Cultural Heritage Applications unit of the Information Society Directorate-General of the European Commission and the Digital Library Initiatives in the United States sponsored by the National Science Foundation and other agencies.

solved. It is based on the primary conclusions of several activities that took place in the context of the DELOS Network of Excellence for Digital Libraries under the Fifth Framework Programme of the European Commission. Among these activities, the brainstorming workshop that was held in San Cassiano [1] played a major role and served as the springboard for many of the subsequent events, including seven working groups [7–13] organised in collaboration with the NSF. Another key activity of DELOS with significant impact was the series of thematic workshops on key digital library research areas [2–6]. For ease of reference, the working groups and thematic workshops are codified as follows:

Working groups:

- G1 *Spoken-Word Digital Audio Collections*
- G2 *Digital Libraries Information-Technology Infrastructures*
- G3 *Personalisation and Recommender Systems in Digital Libraries*
- G4 *ePhilology: Emerging Language Technologies and Rediscovery of the Past*
- G5 *Digital Imaging for Significant Cultural and Historical Materials*
- G6 *Digital Archiving and Preservation*
- G7 *Actors in Digital Libraries*

Thematic workshops:

- W1 *Information Seeking, Searching, and Querying in Digital Libraries*
- W2 *Personalisation and Recommender Systems in Digital Libraries*
- W3 *Interoperability and Mediation in Heterogeneous Digital Libraries*
- W4 *Evaluation of Digital Libraries: Testbeds, Measurements and Metrics*
- W5 *Multimedia Contents in Digital Libraries*

In this paper, the results of the work achieved during the brainstorming and thematic workshops as well as the working group meetings are summarised in a cohesive and comparative way. An analysis of these results appears to point to some key new directions for future digital library research that require a certain change of course with respect to the past. In particular, it is time for *digital library technology* to be developed in its own right and not as an adaptation of techniques borrowed from other disciplines. Diversity of needs among different digital libraries will continue to require specialised approaches to specific problems, but no effort should have to address all aspects of digital library development. The field is ready to design and build generic *digital library management systems (DLMSs)* that will have all the key features that appear fundamental in supporting digital library functionality as it arises in several possible contexts. All specialised functionality should then be developed on top of such systems.

This paper is organised as follows. Section 2 describes a research framework that serves as the foundation for the rest of the paper, including the high-level strategic goals that came out of the main DELOS brainstorming workshop and the main research areas that have been identified as critical for achieving these goals. Sections 3 and 4 present the detailed agendas for each research area as they have been drafted by the participants of the corresponding working groups and thematic workshops, respectively. Section 5 analyses and synthesises the details of these research agendas and reveals certain common patterns. Section 6 uses these patterns to justify in some detail the need for generic DLMSs and charts a high-level roadmap that should be followed to make progress on the research agendas. Finally, Sect. 7 concludes the report and offers some final thoughts on the identity of the digital library field as a whole.

2 Foundational framework

The main brainstorming workshop organised by DELOS produced the so-called San Cassiano Report [1], which formed the foundational framework within which much of the remaining and more detailed research-agenda development took place. In that report, a 10-year Grand Vision for Digital Libraries is specified as follows:

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

In the same report, a conceptual framework for digital libraries is also presented (Fig. 1) that serves as a guide to organising the research required to achieve the above goal. The framework recognises three major layers in a digital library system. At the bottom are the contents of the digital library. On top of these is the core system,

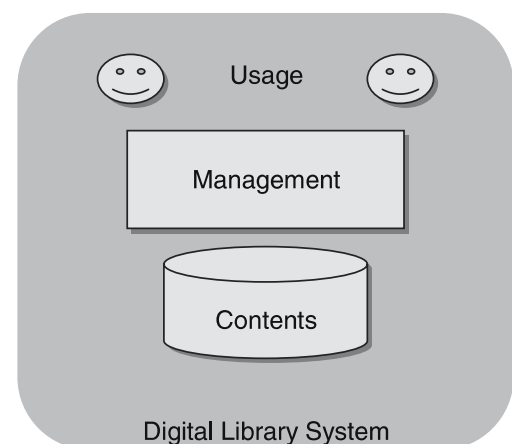


Fig. 1. Conceptual framework for digital library systems

responsible for the management of the contents and for providing the necessary functionality. At the front end is the user interaction layer, dealing with all aspects of the interface between the users and the system.

Note that, on the basis of their primary theme, the DELOS working groups and thematic workshops may be placed within the above framework as follows:

Usage	G3, W1, W2
Management	G2, W3
Contents	G1, G4, G5, G6, W5
Vertical (across all three)	G7, W4

For each of the three main layers of a digital library system, the San Cassiano Report establishes a 10-year goal and analyses the research work that is deemed necessary to reach it. Clearly, any new approach, solution or enhancement in any of these components also affects some or all of the others, generating more related research problems. The report's analysis places within each layer the research problems whose *primary* motivation lies there. These are briefly outlined in the rest of this section.

2.1 Contents

The most critical research issues that arise primarily from the contents layer of a digital library system are the following:

- Collection building, in particular information acquisition, information analysis and extraction, and situated information organisation.
- Collection access and navigation, including efficient search algorithms and structures as well as search optimisation.
- Dealing with non-traditional kinds of objects, such as scientific data collections, simulation models, and combinations of text, video, audio, images, structured data and other forms.
- Dealing with multilingual and multicultural collections, with topics such as culturally driven information translations and management of appropriate linguistic information and meta-information.
- Collection preservation, including software and information migration, translation algorithms and other techniques.

2.2 Management

Likewise, the most critical research issues that arise primarily from the management layer of a digital library system are the following:

- Basic system architecture, with a particular emphasis on component-based, multitier, and Grid architectures.
- Openness, especially working towards flexible systems with plug-and-play functionality, as well as developing various relevant system features, such as auto-description, auto-registration and auto-configuration.

- Interoperability and metadata, such as metadata correlation, registries and conversion tools.
- Scalability, with a special focus on decentralised architectures and performance prediction.
- Availability, in particular advancing the techniques for dynamic reconfiguration and replication.
- Session-flow and workflow management, with modelling, correctness and consistency, long and interoperable sessions, and extensibility being the most central of the relevant issues.
- Security, in particular integrity, confidentiality and digital-rights specification languages.
- Quality, especially establishing quality criteria, metrics, estimation techniques, quality-based processing and quality-oriented metadata.
- Digital library administration.

2.3 Usage

Finally, the most critical research issues that arise primarily from the usage layer of a digital library system are the following:

- User interfaces, including issues of integrated multiparadigm access, task-oriented access and user-interface generation.
- Information visualisation, especially issues related to content-sensitive and context-sensitive visualisation.
- Personalisation and customisation, including explicit and implicit profiling, static and dynamic profiling, personal annotations and person-dependent system behaviour.
- Community information spaces, especially issues regarding implicit and explicit community definition, as well as community annotations and ratings.
- Multilingual and multicultural interactions, emphasising language-dependent and culture-dependent user requests as well as content and service delivery.
- Collaboration, especially synchronous digital library visits.
- Universal access, including all three dimensions of the topic: people (access by everybody), location (access from everywhere) and devices (access via everything, e.g. regular computer screen, palm organiser, etc.).
- Multi-channel access, with particular focus on persistent sessions across multiple devices as well as device-dependent content and service delivery.

2.4 Socio-economic and other issues

In addition to technological research problems, the San Cassiano Report also raises some other important issues that require particular attention in order for the 10-Year Grand Vision to be realised. These are as follows:

- Use of digital libraries in several applications, especially education, including promotion of such use through building digital library services on a variety of educational topics in different parts of the world; stud-

ies on how such use affects learning by various user categories and identification of any particular system demands raised within an educational environment.

- Business modelling, in particular identification of business models for digital library operation and studies of their effectiveness.
- Sustainability, especially mechanisms to fund the continuous renewal of material in digital libraries and maintain users' awareness of their offerings.
- Copyright issues, focusing on possibly identifying a small number of standard approaches that could be used in several cases.
- Methodologies for digital library development, including issues on software, contents collection, daily management and dealing with change.
- Standardisation, with particular emphasis on interoperability.
- Digital libraries as subsystems, in particular methods for interfacing digital library systems with other components of greater systems.

3 The working groups

The DELOS/NSF joint working groups were established in order to explore more deeply some of the research themes that emerged during the San Cassiano Brainstorming Workshop. Some working groups examined a very specific topic while others maintained a relatively more general perspective. In this section, we give a comprehensive summary of the conclusions of the deliberations of each working group. The rest of the papers in this special issue describe in more detail the activities of these working groups and present their recommendations.

3.1 Spoken-word audio collections

Although the number of audio collections is continuously increasing, they are usually inaccessible to the public due to the fact that no adequate search facilities exist and also because they are steadily becoming degraded. Hence, the research agenda charted by this working group focused on providing support for access to and preservation of these audio collections and included, among others, the following issues:

- Analysis and indexing of audio documents according to diverse characteristics, which includes several discrete subtopics: audio partitioning, speech recognition and enhancement, speaker identification and tracking, information extraction and automatic summarisation, and prosodic modelling.
- Browsing and search of audio documents, especially spoken document retrieval, interactive speech retrieval, topic detection and tracking, language identification and cross-language information retrieval.
- Collecting, archiving and preserving audio content, including material filtering, metadata capture, content

reformatting and reshaping, audio-transcript synchronisation, and media storage.

In addition to the above technology-driven issues, the working group also discussed privacy and copyright policies, which are critical topics for the digital library field as a whole.

3.2 Digital library information-technology infrastructure

This working group focused on information systems aspects of digital library technologies and pointed towards *digital library management systems* (DLMSs). It discussed system architecture alternatives with particular emphasis on component-based architectures for scalability and maintainability but also called for investigation of Web services, the Semantic Web, and the Grid. Then, based on two realistic scenarios – one of them dealing with cultural heritage preservation and the other with biomedical research – it specified a five-step user-information cycle and brought out the research issues that are relevant to each step as follows:

- Request specification, such as query languages, new paradigms (e.g. vague personalised queries) and user interfaces.
- Request analysis and evaluation, including new forms of object similarity, multimedia/continuous-media searches, search mechanism combinations, process management and information quality metrics.
- Data storage, from the highest level, where the information space is organised using techniques such as knowledge discovery and data mining, to the level of feature space organisation and multidimensional indexing, on to the lowest level of data format choice.
- Information integration and derivation, with emphasis on multiple result-list combination, visualisation and presentation personalisation.
- Information enrichment, which refers to human post-processing of results and includes the specification of information provenance, general information annotation and document-excerpt combination.

The working group also made recommendations for advancing the supporting environments by developing standard metadata and ontologies.

3.3 Personalisation and recommender systems in digital libraries

Digital library systems must offer content and services that are adapted (in terms of substance and presentation) to the personal needs of the user. These are implied by the user's preferences, task, information needs, background, history, device, location, time of request and other characteristics that essentially define the user's *context*. The relevant research challenges identified by this working group include the following:

- User model design/definition, for flexible and portable models that capture every aspect of the user's context as illustrated above, long-term (persistent) and short-term (ephemeral) models, and individual and group models.
 - User model derivation, implicit methods for learning about the user, collection and analysis of relevant server-side or client-side information for group and individual modelling, respectively, and life-long learning techniques for very long-term modelling.
 - User model processing to personalise system behaviour, identification of the appropriate aspects of the model of the user in each case (e.g., short-term vs. long-term characteristics), and shifting from group-based to individual-based personalisation as the model for a user becomes more detailed.
 - Personalisation of user interaction, personalised interaction styles and information presentation, and trade-off between user control and system initiative.
 - Evaluation of personalised functionality from an individual user or user group perspective, appropriate evaluation criteria and metrics and longitudinal studies for evaluation of very long-term user models.
 - Maintaining balance between personalisation/collaboration and privacy.
- Web/XML information retrieval and the role of different levels of structured markup using languages like SGML and XML.
 - Detection and tracking of entities and relations using a variety of methods, ranging from human-generated rules to more open-ended machine learning.
 - Document understanding of a great variety of forms, such as different levels of summarisation of single or multiple documents, question-answering systems, categorisation and clustering of documents based on their content, filtering and collaborative filtering, and topic detection and tracking.

3.5 *Digital imagery for significant cultural and historical materials*

The key premise for this working group was that digital imagery can be used to protect important cultural and historical materials from perishing and to promote understanding among cultures. Hence, it advocated studying several problems whose solution would facilitate the establishment of sustainable and enduring digital imagery archives of the world heritage. It classified these problems into four areas as follows:

- Digital imagery creation and preservation, including digitisation, (semi-)automatic capture of technical and descriptive metadata, multimodal image acquisition, provenance recording, image compression and storage, and image preservation and restoration, particularly for images of various forms of art.
- Image and video retrieval, emphasising automatic feature extraction and combination for two-dimensional and three-dimensional objects, semantic-based similarity metrics, and search efficiency and scalability.
- Presentation and usability, in particular advanced multimedia user interfaces, new query and browse paradigms, multilingual and language-independent retrieval, multiresolution retrieval, advanced displays, gesture recognition, customisation to user's needs, presentation of three-dimensional objects, and multi-perspective and large-scale visualisation.
- Overall access to the archive, with emphasis on the need to support multiple user groups at the same time, address particularities of different applications, integrate multiple heterogeneous archives, and capture the overall flow of information through the archive.

The working group also gave recommendations on mechanisms for collaborative research, such as creation of a copyright-free corpus of images and videos, development of shared testbeds, and establishment of domain-specific benchmarks.

3.6 *Digital archiving and preservation*

Growing rates of digital content accumulation and the ensuing needs for its retention demand new approaches to digital archiving and long-term preservation. This work-

The working group also emphasised the importance of developing shared platforms (standard data sets, tasks and scenarios) and computational infrastructure to facilitate evaluation of personalisation methods in both laboratory and real-world settings.

3.4 *ePhilology: Emerging language technologies and the rediscovery of the past*

When applied to the vast quantities of historical (mono- or multilingual) material that can be found in museums, libraries and other repositories, emerging language technologies open up several new possibilities to the fields of teaching, learning and research in the broad area of cultural heritage. This working group explored the potential benefits of moving in this direction and emphasised the need to first identify the relevant user communities and understand their attitudes and preferences with respect to information. It also outlined the key technological issues that should be the focus of research in this area, which include the following:

- Evaluation metrics to be used for comparison of the effectiveness and result-quality of language-based systems.
- Visualisation techniques for diverse information that can be extracted from historical and cultural material.
- Information-source quality, especially problems that arise from OCR and the trade-off between having large quantities of low-quality documents versus small quantities of high-quality documents.

ing group analysed the technical challenges that arise in this direction, stressed the need for automated and scalable techniques and recognised the importance of generic solutions whenever possible, such as to interpretability and trustworthiness problems. It identified several research issues that require investigation and partitioned them into three categories, as follows:

- Preservation strategies, such as development of software, format, and device repositories, creation of ‘archival’ media, reconstruction of logical entities from raw data, infrastructure-independent representations and functionality documentation of digital entities, creation of context-aware digital entities, survivability prediction by accelerated ageing and preservation of knowledge.
- Re-engineering preservation processes, including modelling, automation and scalability of processes, trustworthiness and information quality detection, completeness validation of collections and anomaly identification in them, and preservation in distributed and Grid environments.
- Preservation systems and technology, with emphasis on multiple format manipulation, management of complex and dynamic digital entities, automated creation and long-term viability of metadata, use of multilingual entities and technology, appraisal of preservation costs and benefits, and archive repurposing.

In addition, the working group suggested particular research methodologies that could help increase the overall success rate of research efforts in the above areas. It also indicated several non-technical challenges that need to be addressed before preservation research can have the desired impact, with particular emphasis on technology transfer and commercialisation.

3.7 Actors in digital libraries

An actor is any entity (e.g. person, organisation or system) that plays a role in the production, dissemination, management or use of services offered by a digital library. To illustrate the need for research in modelling of actors and roles, this working group introduced an example model consisting of three classes of actors: agents (suppliers of information and services), users (mostly consumers of services) and professionals (managers of digital libraries). Based on this model, the working group analysed the relationships between actor types in the context of four realistic scenarios, each one represented by two cases that were appropriately compared. The first scenario dealt with digital libraries for science and technology, and the comparison was between the traditional competitive approach and an emerging collaborative approach. (More details are presented in the paper included in this special issue that describes the activities of this working group.) The second scenario dealt with discovery services based on harvesting, and the comparison was

between usage of structured and unstructured metadata. The third scenario dealt with libraries for preservation, and the comparison was between conventional collection development and web harvesting. Finally, the fourth scenario dealt with collaborative digital libraries, and the comparison was between closed and open collaborating communities.

This analysis underscored the need for formal and objective reference models for the description of actors and their roles in digital libraries. It showed that the use of a formal model has the potential to reveal patterns of interaction among digital library actors in both practical and theoretical ways, potentially leading to more effective actor-driven applications and more realistic business models. In addition to the need for investigation of reference models, it also brought to light several other issues that warrant further research, such as collection building and preservation, interoperability and metadata, workflow management, security, personalisation and customisation, and interfaces for all users and user groups.

4 The thematic workshops

As mentioned earlier, in parallel with the brainstorming workshops and the working group meetings, the DELOS Network of Excellence also organised several thematic workshops, where the current state of the art and future research directions in particular topics of digital libraries were discussed. In this section, we summarise the ideas expressed by the participants of each thematic workshop regarding the most critical research problems that the field should pursue.

4.1 Information seeking, searching and querying in digital libraries

Retrieval of information from a digital library can take several different forms, including information navigation, searching and querying. Within this broad area, this workshop identified several issues that require attention:

- Combining search approaches and search results, which essentially requires combining evidence for various information pieces for optimal results, with emphasis on efficient algorithms, combining structured and unstructured information, and using XML and RDF.
- Retrieval models and languages, including interaction models that take into account the dynamic nature of digital library search, use of natural language, use of XML, addressing heterogeneity and distribution, and understanding the temporal and spatial dimensions of information.
- Searching multimedia content, especially developing indices that would support different similarity queries on dynamic data for arbitrary metrics, approximate search and distance browsing, as well as content-based

searching, interoperability and personalisation of continuous media.

- System infrastructure, in particular system architecture, functionality, computation support, performance and scalability.

4.2 Personalisation and recommender systems in digital libraries

Earlier work developed a variety of methods that use past user queries or user behaviour to create new queries or recommendations for different media. Nevertheless, the participants of this workshop felt that several open questions remain:

- Creation of algorithms and techniques specific to personalisation needs within digital libraries, including social issues, such as the tradition of privacy in the library science community, which ‘clashes’ with the need for personalisation.
- Evaluation of personalisation and recommendation techniques, which requires the development of rigorous quantitative and qualitative techniques (e.g. with tools from other disciplines like sociology) as well as the establishment of appropriate testbeds.
- Long-term personalisation, in particular devising models that capture the users’ changing needs and interests during time as well as dealing with related interoperability issues that arise when users change or use multiple platforms.

4.3 Interoperability and mediation in heterogeneous digital libraries

Digital collections are inherently heterogeneous and mostly distributed, so they should interoperate through middleware that provides transparent access, data translation, object identification and other related services. Within this framework this workshop raised several issues for future research, broadly classified as follows:

- System architecture, with emphasis on design and generation of mediators, wrappers and agents, scalability, openness to dynamic registration of new collections and efficiency (query processing and optimisation, indexing).
- Metadata, including metadata generation, registration and exchange, metadata categorisation within the Semantic Web, brokering in heterogeneous collections through Web-based end-user systems, use of XML and RDF for service description, and service composition.

4.4 Evaluation of digital libraries: testbeds, measurements and metrics

The methods and metrics used for the evaluation of digital libraries vary according to different perspectives one

may have on digital libraries. Evaluation is different if a digital library is viewed as an institution, as an information system, as a new technology or as a combination of new services. The participants of this workshop emphasised these distinctions as they illustrate the great importance of the following two problem areas:

- Evaluation infrastructure development, in particular creation of an institution that will collect all the records of interaction in digital libraries and incorporate appropriate testbeds that will then be used by the community; it will also provide the necessary infrastructure for subsequent development of common standards.
- Evaluation metrics, especially abandoning current metrics, which have been derived from other contexts, and devising new ones that reflect the totality of digital libraries, as well as developing context-specific and context-sensitive evaluation techniques.

4.5 Multimedia contents in digital libraries

Despite extensive earlier work in the area of multimedia management, this workshop indicated that there are several major unsolved issues that hinder the full incorporation of multimedia items into digital libraries:

- System management, including content addressability, interoperability, use of multiple user-interface devices, scalability and robustness.
- Standard development and use, especially with respect to data and metadata formats, models and languages, with particular emphasis on the Internet and the Web, which should be considered as the standard underlying environment.

5 Research agenda analysis

The previous three sections contain several general and specialised research agendas that have been produced by the various forums that DELOS established for this purpose. In this section, we put the results of all these efforts in the same perspective and draw a general picture of the direction that digital library research should take in the next few years. The conceptual framework introduced in the San Cassiano Report (Fig. 1 and the research issues raised for each of its layers in Sect. 2) serves as the main foundation on which the outputs of the other meetings are placed.² In particular, for each layer of a digital library system, we summarise the subset of its topics that the participants of each working group and thematic workshop examined in depth. At the end of the section, we tabulate all this for easy reference.

² We do not consider non-technological issues (i.e. the socio-economic and other issues listed in Sect. 2) as few activities touched upon any of those.

The *Actors* working group constitutes an exception in that it used a higher-level abstraction of a digital library system and focused on a global issue that is essentially outside the conceptual framework of Fig. 1. It did bring up some of the detailed topics of the framework and addressed them from the perspective of how the different actors and the roles they assume affect them; these are indicated in the table at the end of this section.

5.1 Contents

Because of the nature of digital library systems, with content management being their primary role, most working groups addressed issues of content in some form. The *Spoken-Word* and *Digital Imagery* working groups operated almost exclusively within this part of the conceptual framework, discussing technologies and problems related to acquisition, analysis, access and preservation of audio and imagery collections, respectively. The *Information-Technology Infrastructure* working group addressed information storage and retrieval, data analysis and annotation. The *ePhilology* working group discussed information-source quality, information annotation and markup and document understanding at various levels. Finally, the *Digital Archiving and Preservation* working group targeted long-term collection preservation, which affects all the other contents topics as well.

The same holds for the thematic workshops. The *Information Seeking, Searching, and Querying* workshop examined a variety of topics on collection access and navigation (algorithms and structures) and paid attention to several non-traditional forms of information. The *Interoperability and Mediation* workshop examined various issues of information access and navigation in a heterogeneous and/or distributed environment. The *Evaluation* workshop addressed access and navigation for diverse forms of data, all from the perspective of digital library evaluation. Finally, the *Multimedia Contents* workshop focused on accessing multimedia data in the context of different applications.

5.2 Management

Unlike the case for contents, management was not really the target of many working groups. The *Information-Technology Infrastructure* working group was probably the one that dealt most comprehensively and most generically with the entire spectrum of management issues. The *Digital Preservation and Archiving* working group was also relatively broad in its discussions of management issues, addressed the entire preservation process, emphasised scalability, metadata and information quality, and considered issues of security and distributed architectures. The remaining working groups addressed different issues in a more specific manner. In particular, the *Spoken-Word* working group dealt with metadata for

audio collections and properties like security and quality; the *Personalisation and Recommender Systems* working group mentioned portable user models, workflows for learning user behaviours, and security and quality; the *ePhilology* working group discussed evaluation metrics and briefly touched upon issues of open architectures and metadata for historical material; and the *Digital Imagery* working group focused on interoperability and integration, metadata and workflow management, while it made passing references to security and quality of imagery archives.

Likewise, management issues were the concern of only some of the thematic workshops. The *Information Seeking, Searching, and Querying* workshop paid particular attention to architectural issues, interoperability among heterogeneous digital libraries and several desirable architectural properties, such as scalability and openness. The *Interoperability and Mediation* workshop was the one most intimately related to management. It examined various architectural issues and discussed several desirable system properties for digital libraries. The *Multimedia Contents* workshop also dealt with an extensive set of management issues, focusing particularly, of course, on how these may be addressed for multimedia information. The remaining two workshops had very specific interests in management. The *Personalisation and Recommender Systems* workshop only addressed interoperability and security issues, while the *Evaluation* workshop focused on quality and administration aspects.

5.3 Usage

As with management, most working groups touched upon usage issues only briefly. The *Personalisation and Recommender Systems* working group worked with the full range of usage issues, as personalisation and customisation can affect all of them. The *Information-Technology Infrastructure* working group focused on user interfaces, information visualisation and several new interaction paradigms, including multilingual and personalised. The rest of the working groups addressed some of the issues in a more tangential manner. Thus, the *Spoken-Word* working group discussed briefly multilingual interaction with audio collections as well as advanced presentation of them, the *ePhilology* working group focused on presentation and personalisation of multilingual and multicultural historical material, the *Digital Imagery* working group addressed advanced user interfaces and visualisation, as well as multilingual and multiperspective interaction. Finally, the *Digital Preservation and Archiving* working group discussed only issues on multilingual material.

Likewise, usage issues occupied the participants of only some of the thematic workshops, and for the most part in a relatively limited way. The *Information Seeking, Searching, and Querying* workshop focused only on language aspects of user interfaces and touched upon

issues of personalisation. The *Personalisation and Recommender Systems* workshop explored the corresponding usage topic extensively and also discussed issues of changing platforms. The *Interoperability and Mediation* workshop focused on uniform user interfaces for heterogeneous digital libraries and on collaboration in such an environment. Finally, the *Multimedia Contents* workshop focused on user interfaces and personalisation for multimedia, as well as on how such data can be managed across multiple devices.

5.4 Summary

Table 1 lists the main research issues of the San Casiano Report (Sect. 2) and indicates how they are related to the DELOS working groups and thematic workshops. Clearly, there are many more cells checked in Table 1 than one would expect from the main titles of the working groups and thematic workshops. Apparently, independently of the focus of a particular activity, its participants have invariably identified research problems in other areas as well. This may be interpreted in two ways:

1. By looking at the table column-wise, one may draw the conclusion that, in general, the layers of a digital library system (Fig. 1) are well integrated and interdependent; hence, progress in one requires progress in one or both of the others.
2. By looking at the table row-wise, one sees many topics that appear in the agendas of a large number of the activities and may therefore draw the conclusion that there is room for generic approaches in those topics.

Reconciliation between these two seemingly contradictory conclusions should help to lay the foundations for an overall strategic roadmap that digital library research should follow in order to work towards the 10-Year Grand Vision. This is outlined in the next section.

6 Research roadmap

From the previous research agenda analysis and taking into account the details of the reports of each activity, one can classify the research issues identified into three categories:

1. *Generic issues*: There are several issues that are independent of the others and should be addressed separately. These include several of the items that were particularly popular among the participants of all the activities, e.g. interoperability and metadata, security and quality, as well as some of those items whose nature itself is generic, e.g. system architecture, openness, scalability and availability. Essentially, most (if not all) issues in the management layer of a digital library system belong in this category.
2. *Specialised issues*: By definition, several areas require specialised attention and cannot be affected by progress in others. These mostly include those issues that have been specific targets of a small number of activities, e.g. non-traditional objects, universal access, and multichannel access. Essentially, most issues in the usage layer of a digital library system as well as the really specialised issues in the contents layer belong in this category.

Table 1. Issues addressed by working groups and thematic workshops

Issue	G1	G2	G3	G4	G5	G6	G7	W1	W2	W3	W4	W5
Collection building	✓	✓	✓	✓	✓	✓	✓					
Collection access and navigation	✓	✓		✓	✓	✓		✓		✓	✓	✓
Non-traditional objects	✓			✓	✓	✓		✓			✓	✓
Collection preservation	✓	✓		✓	✓	✓	✓					
Basic system architecture		✓				✓		✓		✓		✓
Openness		✓		✓				✓		✓		✓
Interoperability and metadata	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
Scalability		✓				✓		✓		✓		✓
Availability		✓								✓		✓
Session/workflow management		✓	✓		✓	✓	✓			✓		
Security	✓	✓	✓		✓	✓	✓		✓			
Quality	✓	✓	✓	✓	✓	✓		✓			✓	✓
Digital library administration		✓					✓				✓	
User interfaces		✓	✓		✓		✓	✓		✓		
Information visualisation	✓	✓	✓	✓	✓							✓
Personalisation and customisation		✓	✓	✓	✓		✓	✓	✓			✓
Community information spaces			✓				✓			✓		
Multilingual and multicultural	✓	✓	✓	✓	✓	✓						
Universal access			✓				✓					
Multichannel access			✓						✓			✓

3. *Mixed issues*: Finally, there are issues whose nature is not clean-cut. There is much work that can be done on them in a generic fashion, but there is certainly a limit beyond which progress can only be made by specialised efforts. These issues include those that were discussed in the context of several activities but are not purely generic, e.g. collection building, collection access and navigation, collection preservation, user interfaces, information visualisation, personalisation and customisation, and multilingual and multicultural support. Essentially, most issues in the contents layer of a digital library system as well as some that are in the kernel of the usage layer belong in this category.

If the early myth that all digital library research is by nature specialised is still alive, we believe there is ample evidence in the classification above to put it to rest. The individual research efforts of the past have served the field well by bringing it to its current level of maturity, at the current crossroads. It is now time to change strategy and design and build generic *digital library management systems (DLMSs)* that will have all the key features that appear fundamental in supporting the entire spectrum of digital library functionality as it arises in several possible contexts. In particular, a DLMS should incorporate functionality that is related to all generic issues as well as the generic component of all mixed issues. Moreover, its design should allow for easy integration of sub-systems supporting the specialised functionality required by each particular environment. Figure 2 illustrates this

idea by providing a comparative layout of the research areas in digital libraries and a DLMS.

On a methodological level, the current practice of individual, independent projects, each addressing both generic and specialised issues of digital library support, is inadequate for generic DLMS development. Creation of systems according to the model of Fig. 2 requires focused efforts that will develop generic *digital library technology* in its own right and not as an adaptation of techniques borrowed from other disciplines. By factoring out all issues that appear common to most digital libraries, these efforts will produce the first generation of DLMSs, which will be the main building blocks for all future digital libraries.

Clearly, diversity of needs among different digital libraries will continue to require specialised approaches to specific problems. Hence, the above should be complemented by studies of a host of specialised digital library problems that remain unsolved or for which the current solutions are inadequate, e.g. supporting several forms of rich content or of rich behaviour. This type of work can proceed independently for each problem, but the results should eventually latch onto one or more DLMS designs to give rise to increasingly stronger and richer digital libraries.

7 Conclusions

What has been identified as the 10-Year Grand Vision for digital libraries is certainly far from the current state

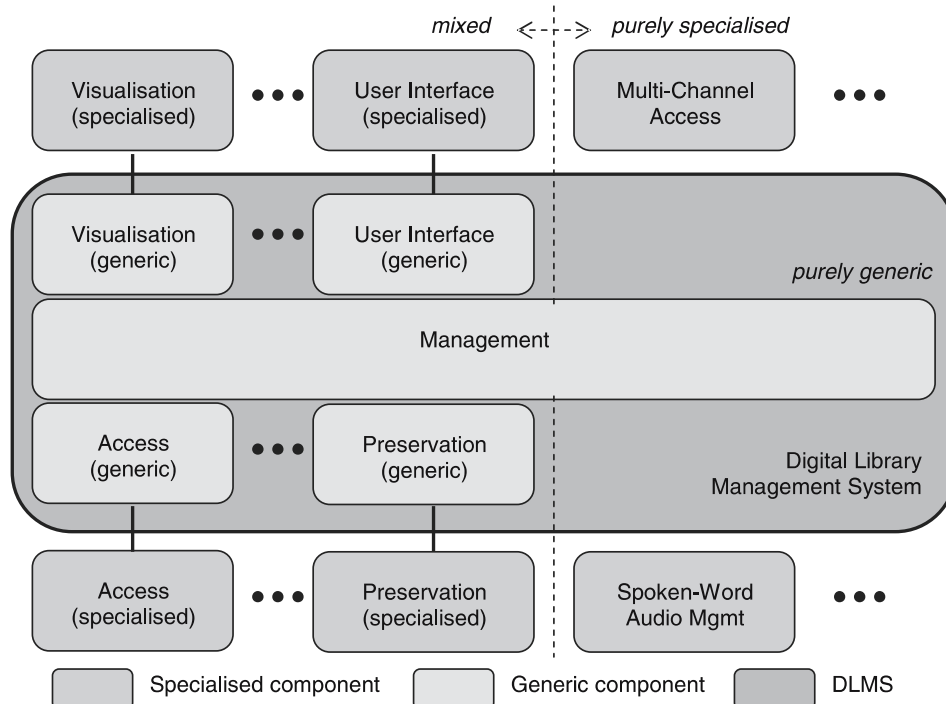


Fig. 2. Conceptual framework for digital library management systems

of the art in the field and may possibly take longer than 10 years to achieve. In fact, its realisation would most likely end the need for any further research, as it would capture more or less the ideal goals in every aspect of the digital library world. Even coming close to this vision, however, will be of tremendous benefit and will imply significant progress on several critical and challenging problems.

In this paper, we call for a concerted research effort aimed at the development of generic digital library management systems. By serving as firm foundations for any further developments, such systems will accelerate advances in all aspects of digital library technology. At the current crossroads of the field, this appears to be the safest, shortest and most fertile way towards the Grand Vision.

Identification of a clear goal and of the path towards it is a major achievement for any scientific community. Having made much progress in this direction, those who serve the field recognise that digital libraries are now moving far beyond their original intentions, far beyond any connotations of the term 'library'.

The potential exists for digital libraries to become the universal knowledge repositories and communication conduits of the future, a common vehicle by which everyone will access, discuss, evaluate and enhance information of all forms [8]. Furthermore, we see the potential for digital libraries to become the strongest shield of humanity protecting its historic, cultural and scientific artefacts from time, natural disasters, thieves, vandals and terrorists [11].

The emerging role of digital libraries now calls for a renaming of the field, as its original name is no longer appropriate. Besides libraries, it also encompasses digital archives and museums; it has moved from capturing just text to dealing with multimedia objects with general knowledge, semantics and behaviour often embedded in them; finally, it supports the specialised needs of very diverse technologies and applications. To capture the essence of its new identity, the field needs a new identification as well. It may be time to move from the **digital library** to the

**Digital Information REALM
(Resources Entrusted to Archives,
Libraries, and Museums)**

Acknowledgements. This work was supported by the Information Society Technologies (IST) Programme of the European Commission as part of the DELOS Network of Excellence (Contract IST-1999-12262). We would like to thank Anya Sotiropoulou and

Dimitris Theotokis for useful material they provided for an earlier version of this paper, Vittore Casarosa, Carol Peters and Costantino Thanos for numerous comments that improved the paper considerably, and Jose Borbigna and Seamus Ross for their suggestions that helped with the accuracy and readability of some of the descriptions in the paper.

References

1. Digital libraries: future directions for a European research programme. DELOS Brainstorming Report, San Cassiano, Italy, June 2001.
<http://delos-noe.iei.pi.cnr.it/activities/researchforum/Brainstorming/brainstorming-report.pdf>
2. Proceedings of the 1st DELOS workshop on information seeking, searching and querying in digital libraries, Zurich, Switzerland, 11–12 December 2000. <http://www.ercim.org/publication/ws-proceedings/DelNoe01/>
3. Proceedings of the 2nd DELOS workshop on personalisation and recommender systems in digital libraries, Dublin City University, Ireland, 18–20 June 2001. <http://www.ercim.org/publication/ws-proceedings/DelNoe02/index.html>
4. Proceedings of the 3rd DELOS workshop on interoperability and mediation in heterogeneous digital libraries, Darmstadt, Germany, 8–9 September 2001. <http://www.ercim.org/publication/ws-proceedings/DelNoe03/index.html>
5. Proceedings of the 4th DELOS workshop on evaluation of digital libraries: testbeds, measurements and metrics, Budapest, Hungary, 6–7 June 2002.
http://www.sztaki.hu/conferences/deval/presentations/DELOSWorkshop4OnEval_report.pdf
6. Proceedings of the 5th DELOS workshop on multimedia contents in digital libraries, Chania, Crete, Greece, 2–3 June 2003. <http://www.music.tuc.gr/MCDL/>
7. Report of the DELOS/NSF working group on spoken-word digital audio collections, 2003.
<http://delos-noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/spokenword/SpokenWord.pdf>
8. Report of the DELOS/NSF working group on digital libraries information-technology infrastructures, January 2003.
<http://delos-noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/DLinfoinfrast/Infrastructures.pdf>
9. Report of the DELOS/NSF working group on personalization and recommender systems in digital libraries, May 2003.
<http://delos-noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/personalization/Personalisation.pdf>
10. Report of the DELOS/NSF working group on ePhilology: emerging language technologies and rediscovery of the past, 2003. <http://delos-noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/ephilology/ephilology.pdf>
11. Report of the DELOS/NSF working group on digital imaging for significant cultural and historical materials, 2003.
<http://delos-noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/digitalimaging/DigitalImaging.pdf>
12. Report of the DELOS/NSF working group on digital archiving and preservation, 2003.
<http://delos-noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/digitalarchiving/Digitalarchiving.pdf>
13. Report of the DELOS/NSF working group on actors in digital libraries, June 2003.
<http://delos-noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/actors/Actors-Roles.pdf>