DELOS
Network of Excellence on Digital Libraries

Future Research Directions

3rd DELOS Brainstorming Workshop Report
Corvara, Italy, 4-6 July, 2004

Editors:

Alberto Del Bimbo – University of Florence, Italy

Stefan Gradmann – University of Hamburg, Germany

Yannis Ioannidis – University of Athens, Greece

DELOS: a Network of Excellence on Digital Libraries

www.delos.info

Information Society Technologies

Project no.507618

DELOS

A Network of Excellence on Digital Libraries

Instrument: Network of Excellence

Thematic Priority: IST-2002-2.3.1.12
Technology-enhanced Learning and Access to Cultural Heritage

D8.1.2 BRAINSTORMING WORKSHOP REPORT

Due date of deliverable: 31 December 2004
Actual submission date: 15 February 2005
Revision date: 8 July 2005

Start Date of Project: 01 January 2004
Duration: 48 Months

Organisation Name of Lead Contractor for this Deliverable: University of Athens

Revision [1]

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)

Dissemination Level: [PU (Public)]

1 Introduction

The vision of Digital Libraries is rapidly changing as a consequence of the advancements and applications of new technological solutions in data communication, storage and retrieval and changes in the modalities in which information can be accessed and retrieved, and at the same time the growth of the maturity of users and the increasing demand for quality and services.

Digital multimedia archives together with technologies to access and present them will provide a resource for both general as well as for specialized activities at any levels. Potential users are educators, students, travelers and members of the general public, with their use running the gamut from curiosity to research to analysis.

Novel technologies or integration of existing technologies should be developed to better facilitate the study and recording of raw elements that have some informative content. These cover system architectures, data acquisition technologies, data representation and storage, retrieval, transmission and presentation. Open standards are needed to facilitate inter-operable systems.

To explore the directions in which Digital Libraries evolve and obtain a fresh perspective on the long-term research agenda of the field, in July 2004 the DELOS Network of Excellence on Digital Libraries organized a brainstorming workshop on "Digital Library Research Directions". A more particular objective was also to outline the main research directions of the future European research programme in the field of Digital Libraries (DLs). For the workshop, DELOS invited approximately 25 prominent researchers both from the Digital Library field as well from some important enabling technologies, two of whom were from the US while the rest from Europe. The workshop was began with an initial plenary session, where the results of the activities of previous DELOS brainstorming workshops, reports from specialized working groups and other similar activities of the FP5 DELOS project were presented. This was followed by presentations of every participant's vision of the future of the field. The next stage involved partitioning of the participants into three separate groups, which met in parallel and were all charged with the task of identifying the critical research issues for the coming decade. In the ensuing plenary session, each group presented its findings, which were discussed by the congregation. The workshop concluded with a first attempt to obtain a synthesis of the three groups' outputs. The present report summarizes the integrated outcome of these parallel efforts, which in general, reflect the results of all of the very intensive discussions that occurred during the two days of the workshop.

The report is organized as follows. Chapter 2 presents the general strategic directions that the field should follow based on the overall conclusions of the workshop. Chapter 3 describes appropriate frameworks at various levels within which future Digital Libraries are expected to operate. Chapter 4 outlines the main topics on which the scientific community is called to conduct research in order to realize the systems envisioned for the future. Chapter 1 touches upon the research methodology that should be followed in addressing the relevant issues. Finally, Chapter 6 summarizes the discussions on a possible new name for the field, to better reflect the new role emerging for Digital Library systems.

2 General Strategic Directions

Looking back at the ten or so years of history of the Digital Libraries field, we observe that almost all efforts to build such systems were characterized by the following features:

- *Content-centric efforts*: Motivation to build Digital Libraries has been based on the needs to organize and provide access to particular collections of data and information.
- *Storage-centric*: The main role of every Digital Library has been assumed to be the static storage and retrieval of information.
- *Environment-specific systems*: Every time, the system was built from scratch and the end product was tailored to the particular needs and characteristics of the target environment.
- *Isolated and repeated efforts*: Primarily as a result of the previous issue, every development effort has been independent of any other and reused nothing already available, hence repeating the work associated with several components that happened to be common.
- *Isolated systems*: Even after development, every Digital Library system has mostly operated in isolation, without any interactions with other similar systems.

Furthermore, so far the Digital Library community has mostly operated under the influence of the following myths:

- Digital Libraries are for things in Libraries only.
- Digital Libraries are for Cultural Heritage only.

Recent proposals for the future of Digital Libraries, however, have given a rather different picture. In particular, the first brainstorming workshop organized by the DELOS Network of Excellence (funding under FP5) produced the so-called `San Cassiano Report' [1], which has formed a foundational framework for some follow-up work in the community. In that report, a 10-year Grand Vision for Digital Libraries has been specified as follows:

> *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.*

In a similar fashion, in the Final Report of DELOS (FP5) [2], a similar in spirit but slightly different 10-year Vision is laid out:

> *The potential exists for Digital Libraries to become the universal knowledge repositories and communication conduits for the future, a common vehicle by which everyone will access, discuss, evaluate, and enhance information of all forms.*

Clearly, any effort towards these two visions requires significant change in the strategies of the past, shifting the focus with respect to functionality and operational environment, and tearing down all myths. In particular, in contrast with the earlier characteristics mentioned above, these two visions call for systems characterized by the following features:

- *Person-centric efforts*: Humans are at the center of Digital Libraries and all efforts must be motivated by needs to provide novel experiences to users.
- *Communication/collaboration-centric*: The main role of Digital Libraries must be to facilitate interaction of scientists, researchers, or the general public on themes that are pertinent to the information stored.
- *Generic technology-based systems*: Generic Digital Library Management Systems (DLMSs) should be developed that capture all common management aspects of

Digital Libraries. Supporting any further, environment-specific needs on content manipulation or user interfaces should be developed in a customized fashion on top of DLMSs.

- *Maximum-reuse efforts*: Assuming the existence of industrial-strength DLMSs, every Digital Library development effort should depend on them, avoiding much mundane work that is currently necessary, and should only focus on the specialized parts.
- *Global distributed systems*: Digital Library systems should be widely interconnected, exchanging and integrating their contents.
- *Universality*: Digital Libraries should be put in the service of ``all'' applications, managing ``all'' forms of content, from data to information to knowledge.

Trying to visualize the differences between the past and the future we refer again to the `San Cassiano Report', where a conceptual framework for Digital Libraries was also presented (**Figure 1**). This served as a guide to organize earlier research agendas and recognized 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, 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.

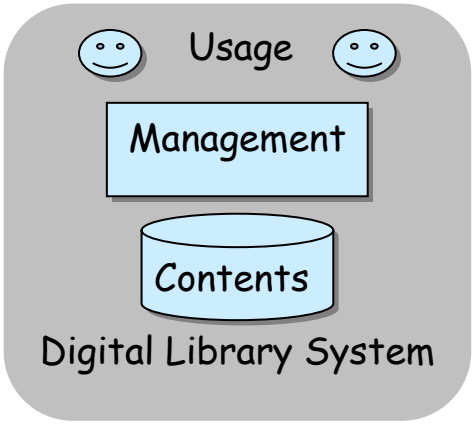


Figure 1: Conceptual Framework for Digital Library Systems

Having the above abstraction in mind, the current universe of Digital Libraries looks like Figure 2. Every system is built starting from the contents and moving up (I→II→III); every system has all of its components different from the others; and they are all isolated.

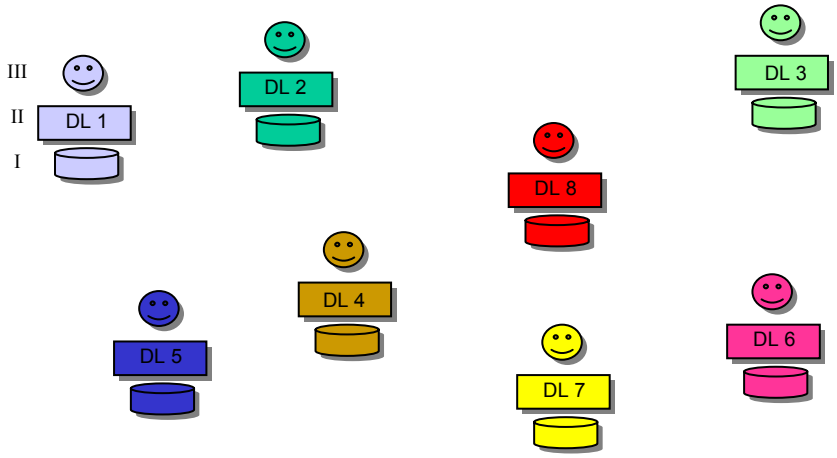


Figure 2: Current Digital Library Universe

Building systems in the way prescribed above, in the direction of the Grand Visions that have been presented by the community will result in Figure 2 being replaced by Figure 3.

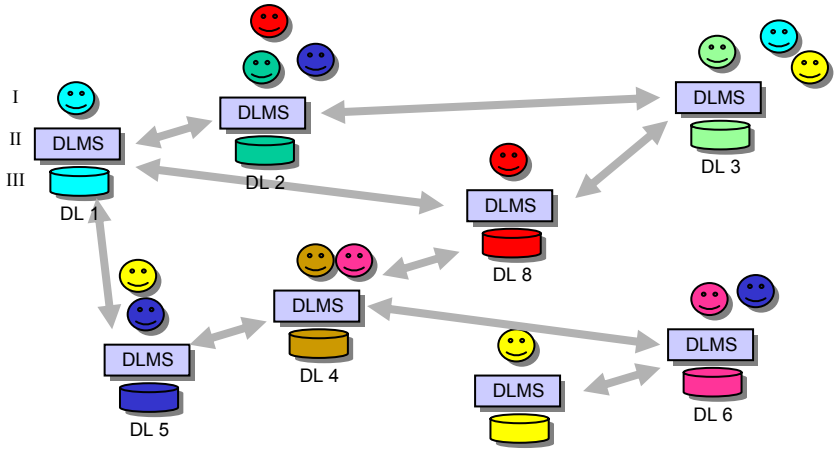


Figure 3: Future Digital Library Universe

Note that every system is built starting from the user needs and moving down (I→II→III); every system has the same, generic DLMS to take care of its basic management functionality and has specialized components only for the rest; all systems are interconnected with each other; and several users interact with each other through this network of systems.

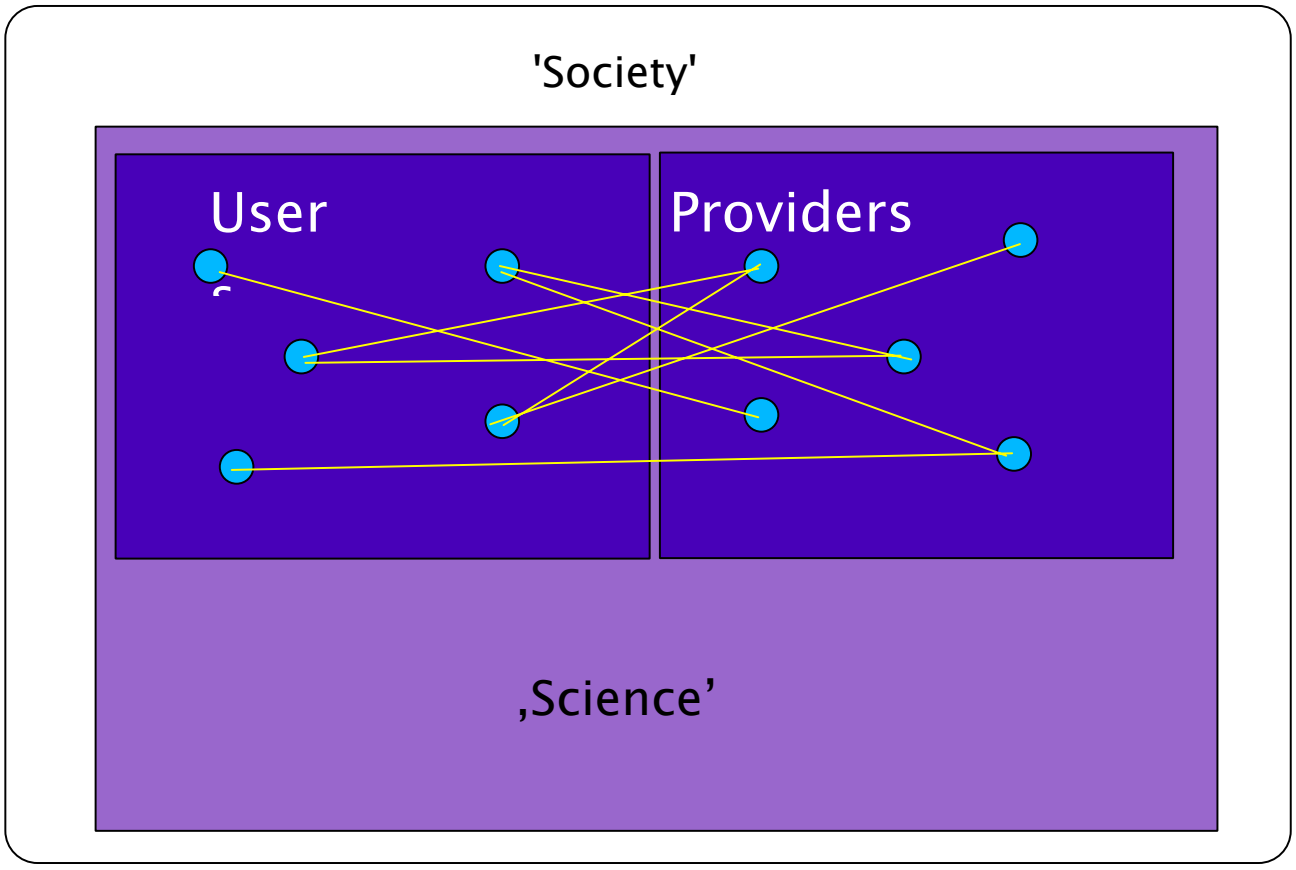
Having the above philosophical and strategic shift in mind, the following sections address in more detail the various conceptual and pragmatic dimensions of the Digital Library area on which the Digital Library community need to reposition itself and the research direction that it needs to explore.

3 Conceptual Frameworks

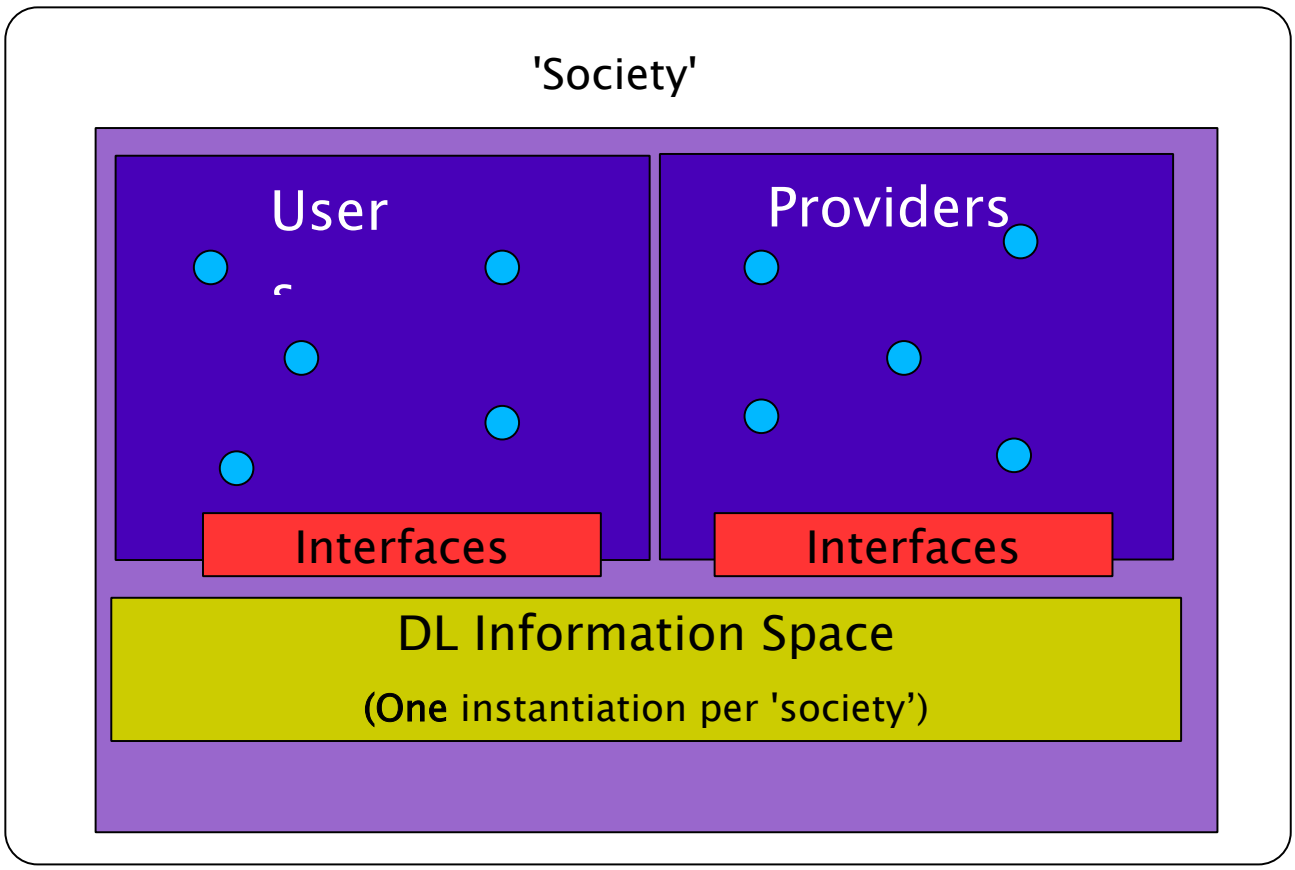
In charting a research agenda, there are several conceptual layers within which one has to look at the object of the research, i.e., Digital Libraries in this case, and present a vision for the future. In the following subsections, we identify several such layers and place next-generation Digital Libraries in the corresponding framework.

3.1 Operational Context

Two contextual levels should be kept in mind when considering scientifically-driven digital information and communication architectures in general as well as Digital Libraries in particular. Within scientific communication contexts (cf. the big blue box in the drawing below) providers and users of scientific information used to establish dedicated, mostly proprietary links for electronic information exchange (the yellow lines linking the turquoise circles). Moreover, the 'science' box and its internal constraints cannot be wholly understood without their societal and cultural context of language, legislation and the like.



One of the primary roles of Digital Library based information architectures is to build a 'Digital Library' information space that would be common for all users and providers of scientific information and which would replace the proprietary communication links with generic, common interfaces to the information space for both users and providers. This target situation is illustrated in the drawing below:



At least in an ideal setting there should be only one instantiation of such a Digital Library information space per given ‘society’ – which in turn is equivalent of a cultural and semiological continuum (not necessarily homogeneous!) among a given group of humans.. The aim thus is not to create a multitude of Digital Libraries but one large, distributed, and still unified information architecture per societal context.

3.2 Stakeholders

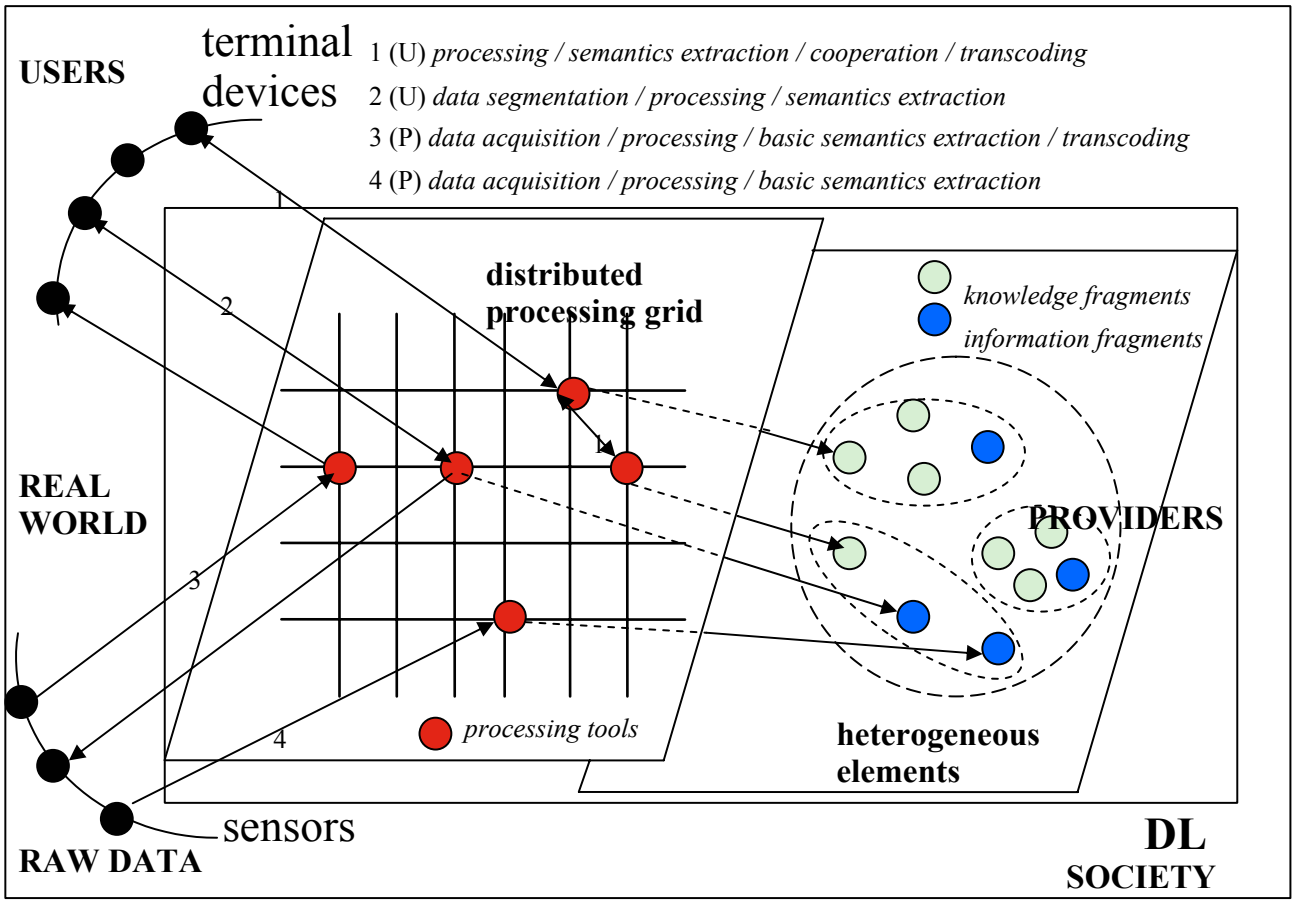
In the traditional societal context, users and providers form two mutually exclusive sets of parties involved in information exchange. In future Digital Libraries, however, these are just different roles that members of a Digital Library environment may play: Users see the Digital Library as a whole and exploit its contents; providers make available information and knowledge fragments that they produced. Both of these roles, however, may be played by the same actor at different times: they are not permanent designations. Such independence between actors and roles repositions Digital Libraries at the center of scientific activity as a critical collaboration and communication tool. Furthermore, it broadens significantly the overall set of beneficiaries of this technology, to include individual subjects, companies, institutions, government organizations, media producers, scientific teams, and several others.

The set of application areas that will benefit from future Digital Libraries has increased as well. Among them, we distinguish health, science, government, culture, and learning, and their `electronic’ counterparts (eHealth and eInclusion, eGovernment, eCulture, and eLearning).

Architectural Environment

The conventional conception of a Digital Library as an isolated repository of particularly-targeted information packaged with intimately associated software belongs to the past. Analogously, the traditional information sources, modes of information creation, and forms of information search are becoming simply alternatives among several options supported

by future Digital Library systems. The following drawing captures a high-level view of the architectural environment anticipated for the coming years as well as some for the most characteristic workflow and information-flow cases that will have to be supported (arrows 1 to 4). The main elements of the latter are outlined in 3.5.



In the above drawing, there are three data abstraction levels that are identified, which are explained below:

- *Raw data*: This is unprocessed data available in different forms and media in the real world
- *Information fragments*: These are results of early processing of raw data and contain some basic semantics
- *Knowledge fragments*: These are results of more advanced processing that is applied on information fragments using user knowledge and context information. They contain higher-level semantics that may tied to the particular application served by the Digital Library environment, such as user-built databases and knowledge units, user models (with specifications of who is, what knows, what is doing at the time, where is, goals and how satisfaction is to be obtained), context models (with specification of location, time, goals), and others.

In the above drawing, one also finds two types of external channels for interaction with the Digital Library system:

- *User terminal devices*: These are devices for human interaction with the system and include new types of displays, such as gesture recognition instruments, etc.
- *Sensors*: These are devices that are deployed in appropriate environments, monitor their surroundings with respect to certain parameters of interest. They feed the system with measurements for processing and storage as well as the users with new information of interest.

Finally, at the center of the drawing above is a `grid' of distributed processing units. Independent of the particular philosophy underlying the distributed architecture (e.g.,

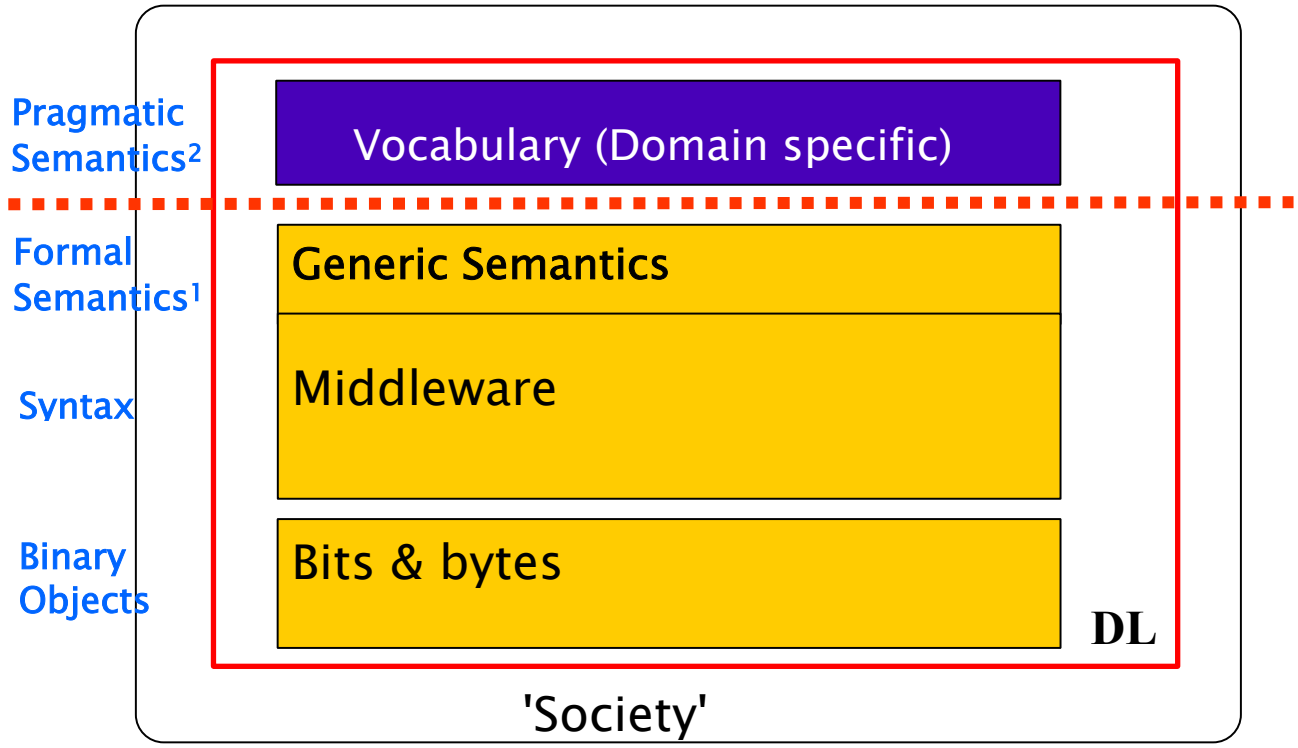
'Grid', peer-to-peer, service-oriented, or any other), the key issue is that processing is an indispensable part of future Digital Library systems, which are populated not just by primary information, but with several products of it as well at various levels and with processing primitives and methods.

3.4 Information Object Abstractions

The drawing below takes up the context related view developed in 3.1 (Operational Context) in order to determine the specific definition realms of Digital Libraries as opposed to the domain specific conception of key scientific information objects. In doing so, it takes as background the Architectural Environment outlined in 0 and tries to identify architectural strata within the Digital Library information model realm as well.

It then appears that a Digital Library information model will certainly include a layer for 'raw' binary data as well as for defining the syntax that can apply to these data elements in order to organize them in a meaningful way – the former being the level of bits and bytes, the latter consisting of middleware components including some of the low level 'semantic web' building blocks which actually are far more syntactic than semantic (e.g., RDF).

The borderline between generic DL-services and domain specific aspects of digital information objects thus resides somewhere within the semantic level. While some very generic semantic operations (basic inference mechanisms and the like) may still fall within the DL realm, domain specific semantics as organized in specialized ontologies definitely are *used*, but not *specified* within the DL context. This methodologically important borderline thus lies somewhere between formal and pragmatic semantics and has consequences in terms of research methodology as we'll see below in section 1 (Research MethodologyResearch Methodology).



3.5 Functional Scope and Workflow Environment

In reference to the drawing of 0, a major future deviation from traditional Digital Libraries is the relative balance between the roles of the system and its users. In particular, all functions provided and supported by the system could be activated by the user directly or indirectly but could also be proactively activated by the system. Pro-activity is realized through spontaneous invocations of processing tools based on user and context profiles. Such pro-activity capabilities will be very important and necessary in the evolving Digital Libraries.

Clearly, the ways work (functions), information, and control may flow through a Digital Library system as depicted in the drawing of 0 are much more complicated and diverse than in today's Digital Libraries. Pro-activity, independence of actors and roles, and the sheer richness of functionality provided are some of the reasons for this increase in complexity. Therefore, capturing and tracking of the workflows associated with all aspects of content management becomes a critical part of Digital Library functionality. This is necessary to ensure that all information processing steps, from creation to dissemination, can be better automated with richer, more intelligible results.

Looking into the functional scope expected from forthcoming Digital Library systems, we identify the following major functional components:

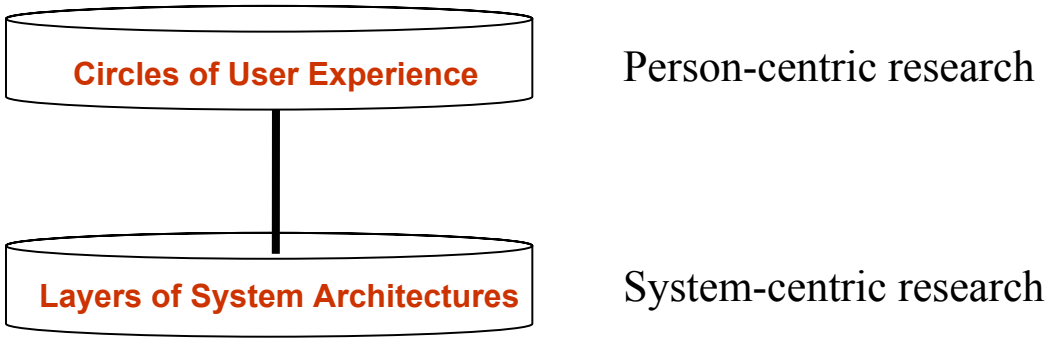
- *Acquisition*: This includes capturing and digitization of physical objects into digital representation; normalization, standardization and cleansing of the captured data; and metadata extraction, both low-level features, e.g., color, textures, shapes, geometries, etc., and high-level features with correlations with other "related" content.
- *Modeling*: This includes creation and update of user profiles; creation and update of contexts of operation; and creation and update of ontologies.
- *Management*: This includes management of both metadata and content (indexing for faster retrieval, addressing data integrity, consistency, and versioning); management of digital identities, ownerships, security and privacy; management of cost policies, quality, trust-ability, and citations; and management of knowledge lifecycle.
- *Mediation*: This includes responsibility to manage heterogeneity and provide information that is consistent to the external environment.
- *Access*: This includes management of data communication, security, context awareness, fair user treatment, and collaborative operations.
- *Distribution*: This includes supporting all the required I/T infrastructures (e.g., caching), copyrights management, etc.
- *Interpretation and translation*: This includes support for languages to represent models, communicate between processing tools, and communicate with the user (natural interaction environments).
- *User-defined task-specific tools*: These include all other pieces of software that will need to be incorporated into the system in order provide customized functionality required by the particular environment or particular `society' targeted by the Digital Library concerned.

4 Research Topics

4.1 Users & Systems

As mentioned in chapter 2, future Digital Library construction is shifting from being content-centric (initiated by the needs of managing content of specific form or meaning) to being person-centric (initiated by the needs of providing specific experiences to users). Digital Library functionality is approached as a variety of desirable user experiences that are layered one within another, from the rudimentary interface-interaction level to the highest cognitive level of exploration and learning. Some of these experiences are extremely novel and unique (e.g., immersive visits to Digital Libraries by multiple users concurrently or personalized annotated conceptualizations of content of interest) while others are relatively well known but nevertheless still challenging in providing (e.g., multi-channel task-specific interactions). Providing these user experiences gives rise to a wide variety of person-centric research problems that require investigation in the years to come.

On the other hand, recent technological advances provide new opportunities for different approaches to Digital Library system development. Furthermore, taking advantage of these opportunities becomes a necessity given the aforementioned user experiences that need to be supported. Several layers of system architectures are required to serve the corresponding levels of data abstractions (3.4) and the corresponding circles of user experiences. Accordingly, devising these system architectures and operating within them gives rise to a wide variety of system-centric research problems that will occupy investigators in the near- and medium-term future.



Across both person-centric and system-centric problems cut several generic, foundational issues that have been troubling Digital Library research since the beginning of the field but have never been actually addressed. As the field has now reached a level of maturity to have its own identity, which permits the development of Digital Library technology in its own right, it is time for some proper foundations to be laid down. Hence, there is a third category of foundational research problems that come to the scene and require attention.

In the following subsections, in a sort of bottom-up fashion, important foundational, system-centric, and person-centric research topics are outlined and briefly analyzed. These are followed by a final subsection, raising awareness to some negative implications that our Digital Library visions may bring upon us and a list of ‘antidote’ research problems that require investigation in order for our worst predictions to never materialize.

4.2 Foundations

Below are listed some foundational issues that are among the most relevant for evolving Digital Libraries.

- *Architectures*: The current typical client-server and 3-tier architectures are not adequate to provide the functionality implied by the advances in all other areas that are relevant to Digital Libraries. Specific effort is needed in exploring novel architectures, particularly the Grid and peer-to-peer approaches, as well as several forms of service architecture.
- *Functionality specification for new objects*: Automatic pattern recognition of raw data can furnish additional annotations relating to physical features of non conventional objects. These automatically generated metadata are distinct from descriptive metadata supplied by experts. For example, still image attributes such as color distribution, shape, texture, and descriptors and invariant descriptors for scale, light, or point of view, are obtained by image analysis. Automatic generation of categories (clustering) of these attributes enables visual overview of discrete image collections. Dynamic video attributes, such as motion field, scene activities, and camera motion, are extracted from the temporal imagery of animation or motion pictures. Other more complex extractable features include automatic transcript generation by speech recognition and geometric 3D-model description.
- *Multi-perspective object access and value tagging*: objects must have different representations (for example still images can be stored together with multiple sets of descriptors and different resolutions) to accomplish different user and application requirements. It will be desirable to monitor the actions of the user so as to add value tags to the data that value reflect the interest that the users have expressed or the context in which the data have been used, or that specialists have preliminarily assigned. The system can accumulate these value tags over the lifetime of the information fragment. This value tagging can be used to enable system to dynamically select between different representations, and narrow searches to specific realms of inquiry.
- *Integrated and unified access to digital repositories*: evolving Digital Libraries must include semantic structures as relationships of concepts allowing high-level content-based retrieval, which have to be integrated with existing retrieval techniques to better facilitate user access. Unified access of multiple archives from different sources is required to support queries across heterogeneous documents. Access of non conventional objects like image, video, sound, music, three-dimensional data ... requires that some machine-based image similarity search is computed by comparing automatically extracted features. Global searches can be narrowed through the use of descriptive metadata.
- *Lifecycle management*: The lifecycle management for a Digital Library includes potentially the capturing and tracking of the workflow associated with all aspects of content so that the processing steps of the content from creation to dissemination can be automated with richer, more intelligible, results. An important consideration is to record the provenance and any subsequent changes/uses of the data. The recording process must incorporate previous use so as to anticipate future needs.
- *Multi-modal functionality combination*: evolving digital collections must support multi-modal (heterogeneous) data handling and their integrated presentation (for example in medical Digital Libraries one should be able to handle photographic images, UV, X-Ray of a patient record in an integrated way). New methods of multi-modal rendering and presentation are also required to support personalization for different audiences and applications. New compression techniques will be needed that support multi-resolution schemes to adapt to the requirements of resolution and transmission speed of a specific application.

- *On-line availability and transcoding*: the on-line extension of digital collections is necessary to share archives among sites of different geographic locations. The technological challenge of online access is to deliver high quality at high speed and low cost to the largest number of users as possible. In presenting query results, the system must dynamically adapt to various devices and bandwidth as well as support personalized formatting of content. Content tagging is needed because it enables us to identify what is important in a given content
- *DLMS specification, design, development* : There is the need of new cost-saving, and effective methods for annotation and retrieval that are easy to use for most users. For example, image analysis techniques provide means to extract useful information from pixels, and provide automatic description of image and video content. These relate to "syntactic" information (like color and texture, video editing effects), low-level primitives (like corners, shapes and spatial relationships) and higher-level information (like objects, scene content, subject description, even associated emotions...) as well as to invariance under different aspects. These annotations can be exploited to perform search of the visual content. Efficient storage and recall of this information to the user is a challenging task.
- *Application interfaces*: Interfaces are tools for building communities (trend to personalized, task-specific, knowledge goal oriented interfaces): We need new interfaces both for non-specialists and for expressing both verbal and non-verbal queries. Such interfaces require:
  - New query and browse paradigms that permit iterative refinement. For non conventional objects, new query types must deal with either low level features (texture, color, shape) or with a high level concept of an item or even metaphors, while operations might include such relationships as logical, temporal, or spatial operations.
  - An abstract layer for posing queries in a way that is independent of data modality and language.
  - A translation layer presenting query results in an intelligible way to the user

  In addition to abstracting and generalizing query paradigms, systems must support both multilingual and language-independent retrieval. This interface challenge is related to the abstract query layer but requires cultural and language dependent ontology management. Since the Internet is currently the most widely used means of information sharing & retrieval, new interfaces must integrate easily with standard Internet browsers (e.g. as plug-ins) and do not require the installation of application-specific search/browsing software.
- *Evaluation (metrics, benchmarks....)*: there is a need to establish benchmark datasets to facilitate evaluation of the research progress. Scalable deployment to archive, search, and analyze very large collections of artifacts is the ultimate challenge. Testbeds may be used among the others, for:
  - Applying new content analysis or CBIR-technology to domain-specific corpus; validating them “at scale”
  - Usage and user studies
  - Evaluation/testing of data exchange and display standards.

4.3 Systems

There is a large number of systemic issues that require investigation before future Digital Libraries can be realized. Some of the most critical ones are the following:

- *System integration, interoperability*: Create complex information systems, that may include designing or building a customized architecture or application, often requires combining components, integrating them with new or existing hardware, packaged and custom. Interoperability requires the ability of two or more systems (or components) to exchange and subsequently use the information they are exchanging. Interoperability requires that a set of implementation standards be identified and implemented by multiple suppliers, and then supported by manufacturers, thus requiring broad industry coordination and collaboration. In this context, semantic interoperability uses explicit semantic descriptions to facilitate information and systems integration.
- *Security, privacy, authorization*: Providing security for digital libraries is challenging due to their dynamic and distributed nature. The need for security in digital libraries arises from legal, social and sensitivity issues of information. Increased dependence on agent-based architectures for digital libraries and peer-to-peer communication makes these more vulnerable to security threats. In addition to the security guarantees, performance and availability guarantees such as convenience in usage, minimal response time, and high throughput are also required of these systems. A secure, precise, standards-based, scalable mechanism thus is needed, enabling information service providers both to guarantee the identity of institutional consumers (authentication) and to provide services based on the characteristics of individual consumers (authorization) while offering increased access, privacy, flexibility and richer management information to consumer institutions.
- *Mobile environments*: Users are allowed to access a Digital Library using different (fixed and mobile) devices at different times. Issues are related to persistent sessions accross multiple devices, and device dependent content and service delivery. In particular, for persistent sessions, 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 their left off. For device-dependent 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 when that is viewed from the same device or a different one with different capabilities.
- *Lossless workflow management*: Workflow engines collect extensive logs of past process executions and make available rich current systems status data. The logs are used for auditing and off-line analysis of the system behavior; the current status data are used by the system administrator to monitor current system behavior. This information can be used for on-line predictions of future behavior of the system and to manage workflows better.
- *Information transformations across media types*: In modern digital libraries, information is encoded in different media types. This often requires that information be transformed from one format to another keeping the same media type, or from one media type to another. In this context, standards, like XML, can easy the task of information transformation accross different media types.
- *Multi-perspective functionality*: To accomplish different user and application requirements, objects can be described at different levels by using different representations and different features. Multiple functionalities can be obtained by

these multi-level features.

- *Optimization of search processes*: Concrete optimization of search processes should involve several aspects, such as search in distributed scenarios, multi-feature and multi-objects search, etc.
- *Service composition, transactional processes, process verification/check*: The need to aggregate or combine small services into larger services is core to a service oriented architecture. In many cases, a single service will act as a front-end to many small services. In fact, a growing trend in software architecture is to build platform-independent software components, like Web services, that are available in the distributed environment of the Internet. Applications are to be assembled from a set of appropriate Web services. Seamless composition of Web services has enormous potential in streamlining business-to-business or enterprise application integration. Transactions cover a significant part of processes on Internet applications and database systems. Beyond the sheer complexity and huge importance of secure, reliable transactions, the difference between a successful and an unsuccessful implementation often rests not just on the big picture, but on the detailed operations of any new systems.
- *Document recombination*: The form in which information is delivered is seldom the most apt for a user's task. A user may obtain information through numerous separate requests and/or in small pieces extracted from multiple documents. Tools should be available that will permit users to interactively segment, combine, restructure, group and organize the individual pieces into new, cohesive documents that capture the users' needs.
- *Information trading in DL federations*: Knowledge sharing and collaboration has always been a hallmark of human activity. Human societies function through cooperation and teamwork: people exchange ideas, information and knowledge to achieve consensus-based decisions. The ubiquity of network connectivity and recent advances in computing technology has raised this interaction to a new level by enabling computer-mediated information sharing and decision-making in all facets of everyday life. Among the application scenarios which now seem feasible are Medical tele-diagnosis, crisis management, mobile office, and electronic trading.
- *Information Preservation:* Preserving information in digital forms is much more difficult than preserving information in other forms, such as paper and film. This is not only a problem for traditional archives, but also for many organizations that have never thought of themselves as performing an archival function.

4.4 User Experiences

Future Digital Libraries will be used in a great variety of novel ways. Supporting them requires extensive research in several areas. Some of the most important ones follow:

- *Balancing cognitive load between user and information environment:* When using a navigation strategy, a user simply has to choose from a menu of destinations. Beyond this, there are no decisions to be made by the reader and no judgement required - simply "click and go". It is important that users remain aware of: *(i)* where they are in the document space; and *(ii)* where they are in their task, and therefore considerable judgment may be required to balance the multiple navigation opportunities against the task priorities. When using a search strategy, the user has to invent a query, evaluate its effectiveness and formulate a refinement strategy. This is potentially a high cognitive load on top of the requirements to keep track of the reading process as outlined above.
- *Collaboration, social context*: Social networking involves people in the creation of all

manner of content: bios, preferences, postings, requests for assistance and direct social networking information, such as social relationships information. Today's social networking solutions have moved toward a closed garden model, where much of this content is inaccessible except by direct navigation, which forms a serious barrier to use. This means that conventional search engines can't spider the social network content. As problematic as the issues surrounding social content are, the issue of social context is significantly worse. The social context problem centers on a person's heterogeneous social network; being in fact a collection of independent networks, whose purpose and ethos are defined by context. Without the ability to partition the network — based on interests, affiliation and/or reputation — the answers will start to look like Google. The answer is that what is really needed is to pose social context-based questions. Until social networks attack this angle, we will be dealing with a very coarse-grained approximation for what is actually happening in social interactions.

- *Personalized, customized, context-dependent user/information interaction:* The term context is mainly associated to the concept of location, but is far richer than that; different categories of context, such as computational, user, physical and temporal context can be also considered.
- *Harvest information across linguistic boundaries*: 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 systems behaviour is significant and much further reaching than that of other type 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.
- *Adding personal memory to global memory, originator notification*: Memory from people from the past and the present can now be shared to form a sort of global memory derived as the summation of personal, individual memory and knowledge.
- *Distributed (peer-to-peer) data creation/publication:* In a Peer-To-Peer (P2P) system, autonomous computers pool their resources (e.g., files, storage, compute cycles) in order to inexpensively handle tasks that would normally require large costly servers. The scale of these systems, their "open nature," and the lack of centralized control, pose difficult performance, search and security challenges.
- *Communication via annotations:* The widespread opportunities for users themselves to modify web data via annotation have inaugurated a completely novel form of communication. Many frequently used web-sites allow users to comment on or review the data they support. For example, you may be able to add a review of a movie or comment on your experiences operating a tool that is being advertised. You are able to *annotate* their data, and other people find your annotations when they look at the same data. Sometimes you can “annotate annotations” by providing comments on other reviews. Another form of annotation is to be found in the management of e-mail “threads”. As a consequence, annotation is becoming a new form of communication, and that understanding and managing this new medium presents a major challenge.
- *Accessibility, and usability*: As the Web evolves, new software and applications for use on Websites are being developed. Many of these new applications are proprietary products that don't use standard features recognised by the World Wide Web Consortium (W3C). The use of non-standard formats can cause significant accessibility problems for some people. Where it is not possible to use a W3C technology, or doing so results in material that does not transform gracefully, alternative versions of the content that is accessible must be provided. Similarly,

usability issues arise for visual interfaces of new visualization and access devices.

- *New display solutions*: Especially large-scale displays, which have been shown to generate different impact on users and enable greater understanding of complex data.
- *Seamless interfaces*: A seamless interface joins two computer programs so that they appear to be one program with a single user interface. Seamless interfaces can integrate information about an entire collection with recent queries against the collection.
- *Interaction based on gesture recognition*: May include eye, hand, head, position movement, attention span tracking and even mood (based on brain wave detection). The collection of such gestures may also help to create a natural way to interact with the viewer by permitting the simulation of virtual guides.
- *Continuous personalization of a common interface*: Using continuous personalization, the same interface can be adjusted for different user roles using various devices.

4.5 Vision Danger Antidotes

The vision for future Digital Libraries that has been described in this document has as its center piece a fully-distributed, pervasive, ubiquitous storage and computing system. This becomes the one most important tool in the hands of users for their information and processing needs, while by being pro-active, it also becomes to a large extent the leader to information discovery. Such tight dependence of users on a system may be a source of several potential problems at the cognitive and sociological level. Among them we distinguish the following three (*to be expanded*):

- Dealing with the avalanche of data generated within the environment of future Digital Libraries
- Protection of individuals' privacy from the multitude of sensors that will be planted in wide areas for information gathering (e.g., intelligent dust)
- Controlling individuals' profile creation to maintain the appropriate balance between protecting users' privacy and offering personalized services

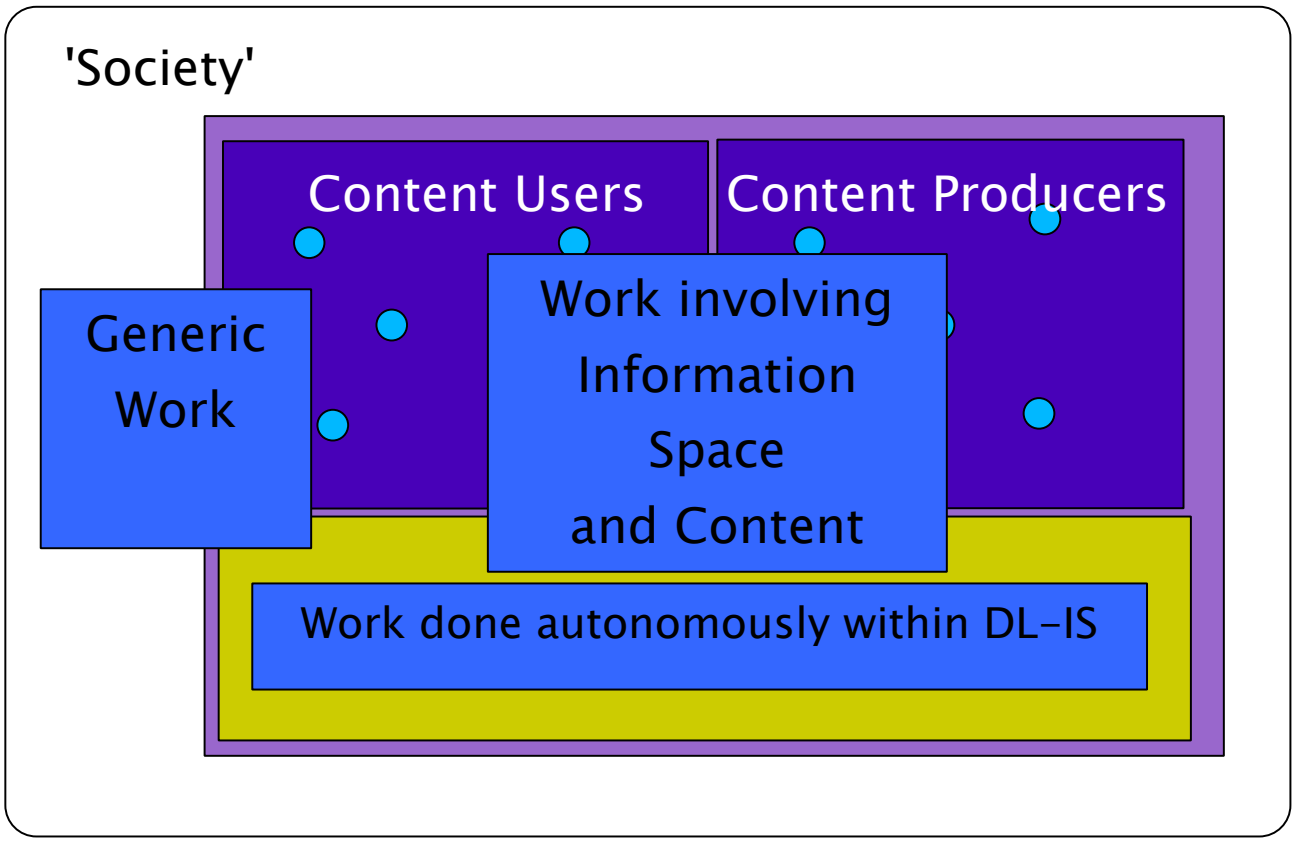
In order for the Digital Library vision to be successful and accepted when it is realized, research should be conducted and solutions should be found on the above problems.

5 Research Methodology

Taking up reflection elements from 3.1 (Operational Context) and 3.4 (Information Object Abstractions) one can conclude that on a methodological level there are three distinct layers of research that should be kept apart in order not to create unnecessary confusion:

- Substantial work can be done autonomously within the DL information space, without involving outside parties
- Research in other areas, however, associated with substantial parts of pragmatic, domain-specific semantics, can only be done involving both DL and content oriented communities. It is crucial to create settings that bring these two worlds together whenever this is required instead of trying to do this kind of work autonomously within the DL information space.
- Still other areas require the participation of societal or cultural instances to such a degree that DL specific considerations can only be made as part of larger, highly generic co-operation scenarios, in which DL actors mostly will not be the driving force.

The distinction is illustrated in the drawing below: assigning the correct methodological level to DL related research issues as listed above probably is one of the crucial managerial tasks in steering the DL research agenda.



6 “What’s in a name?”

Concluding this report, we bring out the last part of the discussion that occurred among the participants of the DELOS Brainstorming Workshop in Corvara. It was proposed that the emerging new role of Digital Libraries may be calling for a renaming of the field, as its original name may no longer be appropriate. Studying various alternatives that were proposed revealed that all of them followed the pattern:

(adjective) (content abstraction) (created entity)

For example, the current name `Digital Libraries’ follows this pattern by the following assignment to the pattern components: (adjective) = ‘digital’, (content abstraction) = <>, and (created entity) = ‘library’. The various proposals for a new name contained several thoughts about each part, besides having nothing (<>) for a particular part:

(adjective)	digital, pervasive, ubiquitous, dynamic, ambient, collaborative, universal
(content abstraction)	data, information, knowledge
(created entity)	Libraries, environments, spaces, gardens, architectures, forums, factories, realms, agoras, worlds

An informal vote among the workshop’s participants brought out

`*Dynamic Universal Knowledge Environments*’

(DUKEs) as the most representative name to express future systems with the kind of functionality outlined in this report. Whether or not this will be adopted by the community remains to be seen. Moving forward with the outlined research, however, is the highest priority.

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. "Digital Libraries at a Crossroads", DELOS Final Report, July 2003 (also to be published in the Journal of Digital Libraries)