D-Lib Magazine
March/April 2007

Volume 13 Number 3/4

ISSN 1082-9873

Setting the Foundations of Digital Libraries

The DELOS Manifesto

Leonardo Candela, Donatella Castelli, Pasquale Pagano, Constantino Thanos
Institute of Information Science and Technologies (ISTI)
Italian National Research Council (CNR)
Pisa, Italy

Yannis Ioannidis, Georgia Koutrika
Department of Informatics and Telecommunications
University of Athens
Athens, Greece

Seamus Ross
Humanities Advanced Technology and Information Institute (HATII)
University of Glasgow
Glasgow, United Kingdom

Hans-Jörg Schek
Department of Computer and Information Science
University of Konstanz
Konstanz, Germany

Heiko Schuldt
Database and Information Systems Group
University of Basel
Basel, Switzerland

(The Point of Contact for this article is <Donatella.Castelli@isti.cnr.it>.)

Abstract

The term "Digital Libraries" corresponds to a very complex notion with several diverse aspects and cannot be captured by a simple definition. A robust model of Digital Libraries encapsulating the richness of these perspectives is required. This need has led to the drafting of *The Digital Library Manifesto*, the aim of which is to set the foundations and identify the cornerstone concepts within the universe of Digital Libraries, facilitating the integration of research results and proposing better ways of developing appropriate systems. The *Manifesto* is a result of the collaborative work of members of the European Union co-funded DELOS Network of Excellence on Digital Libraries.[1] It exploits the collective understanding that has been acquired, over more than a decade, on Digital Libraries by European research groups active in the Digital Library field, both within DELOS and outside, as well as by other groups around the world.

This article presents the core parts of the *Manifesto* that introduce the entities of discourse of the Digital Library universe.

1 Introduction

Digital Libraries represent the meeting point of many disciplines and fields, including data management, information retrieval, library sciences, document management, information systems, the web, image processing, artificial intelligence, human-computer interaction, and digital curation. This multidisciplinary nature has led to a variety of definitions as to what a Digital Library is, each one influenced by the perspective of the primary discipline of their proposer(s) [7][10][11][12][16][2]. The concept of Digital Library has evolved substantially since the early idea of it as a system for providing access to digitized books and other text documents. The DELOS Network of Excellence on Digital Libraries [4], for example, now envisions a Digital Library as a tool at the center of intellectual activity having no logical, conceptual, physical, temporal, or personal borders or barriers to information [5]. Generally accepted conceptions have shifted from a content-centric system that merely supports the organization and provision of access to particular collections of data and information, to a person-centric system that delivers innovative, evolving, and personalized services to users. Conceptions of the role of Digital Libraries have shifted from static storage and retrieval of information to facilitation of communication, collaboration, and other forms of dynamic interaction among scientists, researchers, or the general public on themes that are pertinent to the information stored in the Digital Library. Moreover, expectations of the capabilities of Digital Libraries have evolved from handling mostly centrally located text to synthesizing distributed multimedia document collections, sensor data, mobile information, and pervasive computing services.

This vision of Digital Libraries seems to resonate well with the concept of "Information Space" as established within the field of Computer Supported Cooperative Work (CSCW). Snowdon, Churchill, and Frecon [15] have developed future visions about "Connected Communities" and "Inhabited Information Spaces", with the latter being closely related to the vision of Digital Libraries, in that ubiquitous information is a prerequisite for CSCW.

The variety of conceptions of what a Digital Library is has had a substantive impact on attempts to define and bound the term 'Digital Library'. Since 2006 the term has been generally used to refer to systems that are heterogeneous in scope and provide diverse types of functionality. These systems include digital object and metadata repositories, reference-linking systems, archives, content administration systems (mainly developed by industry), and complex systems that integrate advanced digital library services (mainly developed in research environments). This "overloading" of the term 'Digital Library' results in Digital Library services and systems that do not deliver interoperability and reuse of content and technologies.

Given the current level of maturity and substantial knowledge and experience that have been accumulated in recent years, the time is certainly ripe to lay the necessary foundations for overcoming these challenges by facilitating integration of relevant research results and improving system development methodologies. In order to respond to this need, and following the example of other IT research areas,[2] the members of the DELOS Network of Excellence on Digital Libraries have published the *Digital Library Manifesto*. The *Manifesto* establishes principles to underpin the field and lead to the development of reference documents that will capture the full spectrum of concepts playing a role in Digital Libraries. The *Manifesto* exploits the collective understanding of Digital Libraries developed by European research groups, including those that are partners in DELOS, and the results of DELOS working meetings (e.g., San Cassiano in 2001, Corvara in 2004, and Frascati in 2006).

In this article, we present the core elements of the *Manifesto* and introduce central aspects of the Digital Library framework. The discussion begins with an examination of the three types of relevant "systems" in this area: Digital Library, Digital Library System, and Digital Library Management System. It also explains how they interrelate. The discussion then moves on to

examine three other core topics: the key concepts that characterize these systems encompassing content, user, functionality, quality, policy, and architecture; the range of roles that actors play in digital libraries from application developer to administrator, to designer, and finally to end-user; and clarification of the different levels of abstraction that help us to talk intelligently about the DL Universe.

2 The Digital Libraries Universe: A Three-tier Framework

A Digital Library is an evolving organization that comes to existence through a series of development steps that bring together all necessary constituents. Figure 1 presents this process pictorially and indicates three distinct notions of "systems" developed along the way: Digital Library, Digital Library System, and Digital Library Management System. These correspond to three different levels of conceptualization of the universe of Digital Libraries.

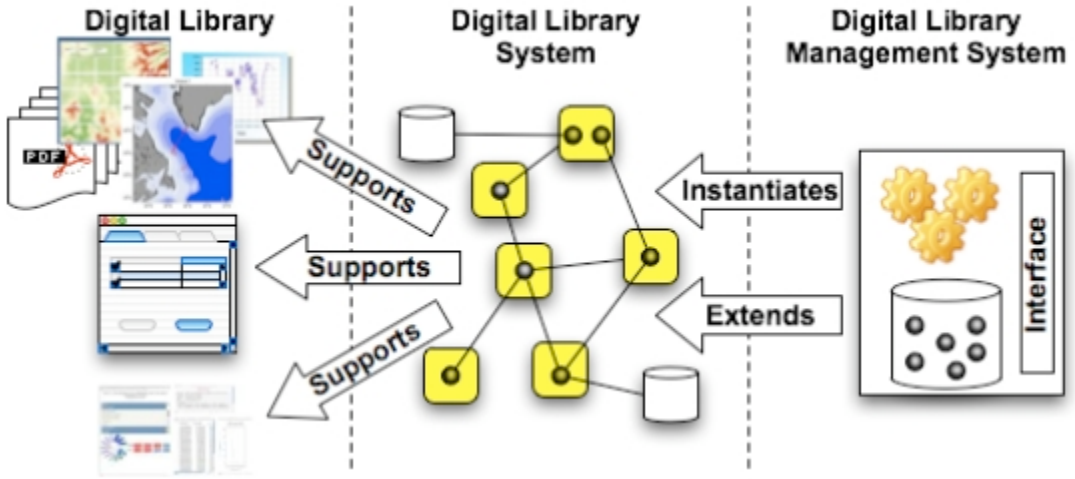

Figure 1. DL, DLS, and DLMS: A Three-tier Framework

These three system notions are often confused and are used interchangeably in the literature; this terminological imprecision has produced a plethora of heterogeneous entities and contributes to making difficult the description, understanding and development of digital library systems. As Figure 1 indicates, all three systems play a central and distinct role in the Digital Library development process. To clarify their differences and their individual characteristics, the explicit definitions that follow may help:

Digital Library (DL)
*A possibly virtual organization that comprehensively collects, manages, and preserves for the long term rich **digital content**, and offers to its **user** communities specialized **functionality** on that content, of measurable **quality** and according to codified **policies**.*

Digital Library System (DLS)
*A software system that is based on a defined (possibly distributed) **architecture** and provides all functionality required by a particular Digital Library. Users interact with a Digital Library through the corresponding Digital Library System.*

Digital Library Management System (DLMS) *A generic software system that provides the appropriate software infrastructure both (i) to produce and administer a Digital Library System incorporating the suite of functionality considered foundational for Digital Libraries and (ii) to integrate additional software offering more refined, specialized, or advanced functionality.*

A Digital Library Management System belongs to the class of "system software". As is the case in other related domains, such as operating systems, databases, and user interfaces,

DLMS software generation environments may provide mechanisms to be used as a platform to produce Digital Library Systems. Depending on the philosophy it follows, a DLMS belongs to one of the following three types:

- **Extensible Digital Library System**
  A complete Digital Library System that is fully operational with respect to a defined core suite of functionality. DLs are constructed by instantiating the DLMS and thus obtaining the DLS. Thanks to the open software architecture, new software components providing additional capabilities can be easily integrated. The DelosDLMS [14] is a prototypical example of a system based on this philosophy.
- **Digital Library System Warehouse**
  A collection of software components that encapsulate the core suite of DL functionality and a set of tools that can be used to combine these components in a variety of ways (in Lego®-like fashion) to create Digital Library Systems offering a tailored integration of functionalities. New software components can easily be incorporated into the Warehouse for subsequent combination with those already there. BRICKS [3] and DILIGENT [6] are two prototypical examples of systems that are based on this philosophy.
- **Digital Library System Generator**
  A highly parameterized software system that encapsulates templates covering a broad range of functionalities, including a defined core suite of DL functionality as well as any advanced functionality that has been deemed appropriate to meet the needs of the specific application domain. Through an initialization session, the appropriate parameters are set and configured; at the end of that session, an application is automatically generated, and this constitutes the Digital Library System ready for installation and deployment. The MARIAN framework equipped with the 5SL specification language represents an example of this process [9].

Although the concept of Digital Library is intended to capture an abstract system that consists of both physical and virtual components, the Digital Library System and the Digital Library Management System capture concrete software systems. For every Digital Library, there is a unique Digital Library System in operation (possibly consisting of many interconnected smaller Digital Library Systems), whereas all Digital Library Systems are based on a handful of Digital Library Management Systems.[3] For instance, through DILIGENT it is possible to build and run a number of DLSs, each realising a DL serving a target community. The DL is thus the abstract entity that "lives" thanks to the software system constituting the DLS.

3 The Digital Libraries Universe: Main Concepts

Despite the seeming richness and diversity of existing digital libraries,[4] in actuality, there is only a small number of core concepts defined by all systems. These concepts are identifiable in nearly every Digital Library currently in use. They serve as a starting point for any researcher who wants to study and understand the field, for any system designer and developer intending to construct a Digital Library, and for any content provider seeking to expose its content via digital library technologies. In this section, we identify these concepts and briefly discuss them.

Six core concepts provide a foundation for Digital Libraries. Five of them appear in the definition of Digital Library: *Content*, *User*, *Functionality*, *Quality*, and *Policy*; the sixth one emerges in the definition of Digital Library System: *Architecture*. All six concepts influence the Digital Library framework, as shown in Figure 2.

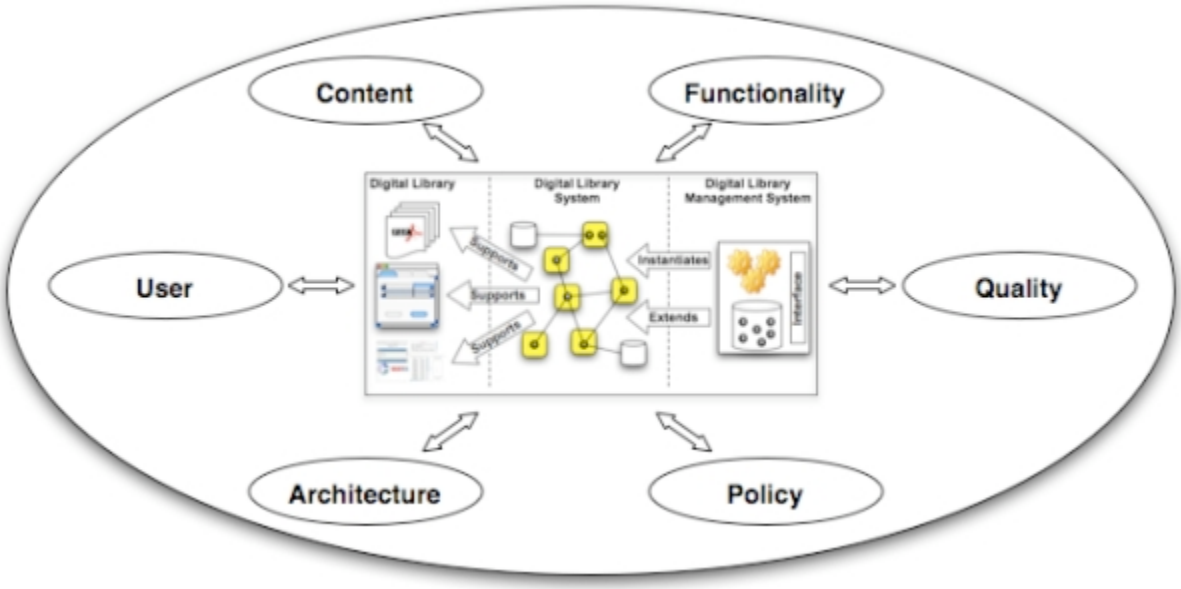

Figure 2. The Digital Library Universe: Main Concepts

Content
The *Content* concept encompasses the data and information that the Digital Library handles and makes available to its users. It is composed of a set of information objects organized in collections. Content is an umbrella concept used to aggregate all forms of information objects that a Digital Library collects, manages, and delivers, and includes primary objects, annotations, and metadata. For example, metadata have a central role in the handling and use of information objects, as they provide information critical to its syntactic, semantic, and contextual interpretation.

User
The *User* concept covers the various actors (whether human or machine) entitled to interact with Digital Libraries. Digital Libraries connect actors with information and support them in their ability to consume and make creative use of it to generate new information. User is an umbrella concept including all notions related to the representation and management of actor entities within a Digital Library. It encompasses such elements as the rights that actors have within the system and the profiles of the actors with characteristics that personalize the system's behavior or represent these actors in collaborations.

Functionality
The *Functionality* concept encapsulates the services that a Digital Library offers to its different users, whether classes of users or individual users. While the general expectation is that DLs will be rich in capabilities and services, the bare minimum of functions would include such examples as new information object registration, search, and browse. Beyond that, the system seeks to manage the functions of the Digital Library so that they reflect the particular needs of the digital library's community of users and/or the specific requirements relating to the information resources it contains.

Quality
The *Quality* concept represents the parameters that can be used to characterize and evaluate the content and behavior of a Digital Library. Quality can be associated not only with each class of content or functionality but also with specific information objects or services. Some of these parameters are objective in nature and can be automatically measured, whereas others are subjective in nature and can only be measured through user evaluations (e.g., focus groups).

Policy

The *Policy* concept represents the set or sets of conditions, rules, terms and regulations governing interaction between the Digital Library and users, whether virtual or real. Examples of policies include acceptable user behaviour, digital rights management, privacy and confidentiality, charges to users, and collection delivery. Policies belong to different classes; for instance, not all policies are defined within the DL or the organization managing it. The policy concept supports the distinction between extrinsic and intrinsic policies. The definition of new policies and re-definition of older policies will be a feature of Digital Libraries.

Architecture
The *Architecture* concept refers to the Digital Library System entity and represents a mapping of the functionality and content offered by a Digital Library onto hardware and software components.[5] There are two primary reasons for having Architecture as a core concept: (i) Digital Libraries are often assumed to be among the most complex and advanced forms of information systems [8]; and (ii) interoperability across Digital Libraries is recognized as a substantial research challenge. A clear architectural framework for the Digital Library System offers ammunition in addressing both these issues effectively.

The concepts of Content, User, Functionality, and Policy share many similar characteristics and are all concepts referring to internal entities of a Digital Library that can be sensed by the external world. Introducing the *Resource* concept, which subsumes the above four, enables us to reason about these characteristics in a consistent manner.

The six core concepts (Content, User, Functionality, Quality, Policy and Architecture) that lie at the heart of DLs need to be considered in conjunction with the four main ways that actors interact with digital library systems, as discussed in the next section.

4 The Digital Library Universe: The Four Main Roles of Actors

We envisage actors interacting with Digital Library Systems playing four different and complementary roles: *DL End-Users*, *DL Designers*, *DL System Administrators*, and *DL Application Developers*.

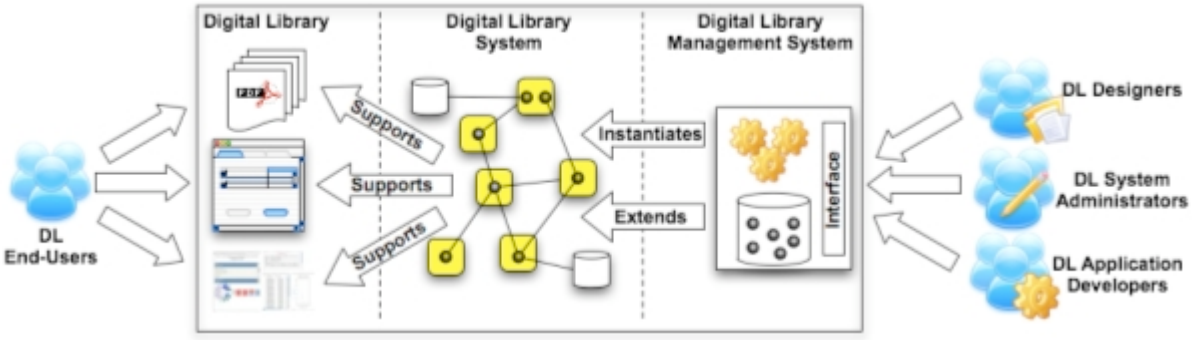

Figure 3. The Main Roles of Actors versus the Three-tier Framework

As shown in Figure 3, each role is primarily associated with one of the three "systems" in the three-tier framework.

DL End-Users
DL End-Users exploit the DL functionality for the purpose of providing, consuming, and managing the DL Content and some of its other constituents. They perceive the DL as a stateful entity serving their functional needs. The behavior and output of the DL depend on the DL's state at the time a particular part of its functionality is activated. The state of the DL corresponds to the state of its resources, which as we have seen above, consists of the collections of information objects managed by the DL, the set of authorized users, the DL's functionality, and its set of policies. This state changes during the lifetime of the Digital Library according to the functionality activated by users and their inputs. DL End-Users may be further

partitioned into *Information Creators*, *Information Consumers*, and *Librarians*.

DL Designers
DL Designers exploit their knowledge on the semantics of the application domain in order to define, customize, and maintain the Digital Library so that it is aligned with the information and functional needs of its potential DL End-Users. To perform this task, they interact with the DLMS providing functional and content configuration parameters. Functional parameters instantiate aspects of the DL functionality that are to be perceived by the DL End-Users, including the characteristics of the result set format, query language(s), user profile formats, and document/data model employed. Content configuration parameters specify third-party resources exploited by the specific DL, e.g., repositories of content, ontologies, classification schemas, authority files, and gazetteers. The values of these parameters configure the way the DL will be presented to the DL End-Users, because they determine the particular Digital Library System instance serving the Digital Library. Of course, these parameters need not necessarily be fixed for the entire lifetime of the DL; they may be reconfigured to enable the DL to respond to the evolving expectations of users and changes in all aspects from policies to content.

DL System Administrators
DL System Administrators select the software components necessary to construct the Digital Library System. Their choice of elements reflects the expectations that DL End-Users and DL Designers have for the Digital Library, as well as the requirements that the available resources impose on the definition of the DL. DL System Administrators interact with the DLMS by providing architectural configuration parameters, such as the chosen software components and the selected hosting nodes. Their task is to identify the architectural configuration that best fits the DLS in order to ensure the highest level of quality. The values of the architectural configuration parameters can be changed over the DL lifetime. Changes of configuration parameters may result in the provision of different DL functionality and/or different quality levels.

DL Application Developers
DL Application Developers develop the software components of DLMSs and DLSs, to ensure that the appropriate levels and types of functionality are available.

The four roles described above encompass the whole spectrum of actors interacting with digital libraries. Their models of the DL Universe are linked together in a hierarchical fashion, as depicted in Figure 4. This hierarchy is a direct consequence of the above definitions, since DL End-Users act on the Digital Library, whereas DL Designers, DL System Administrators and DL Application Developers operate on the DLMS and, consequently, on the DLS and DL as well. This inclusion relationship ensures that co-operating actors share a common vocabulary and knowledge. For instance, the DL End-User expresses requirements in terms of the DL model and, subsequently, the DL Designer understands these requirements and defines the DL accordingly.

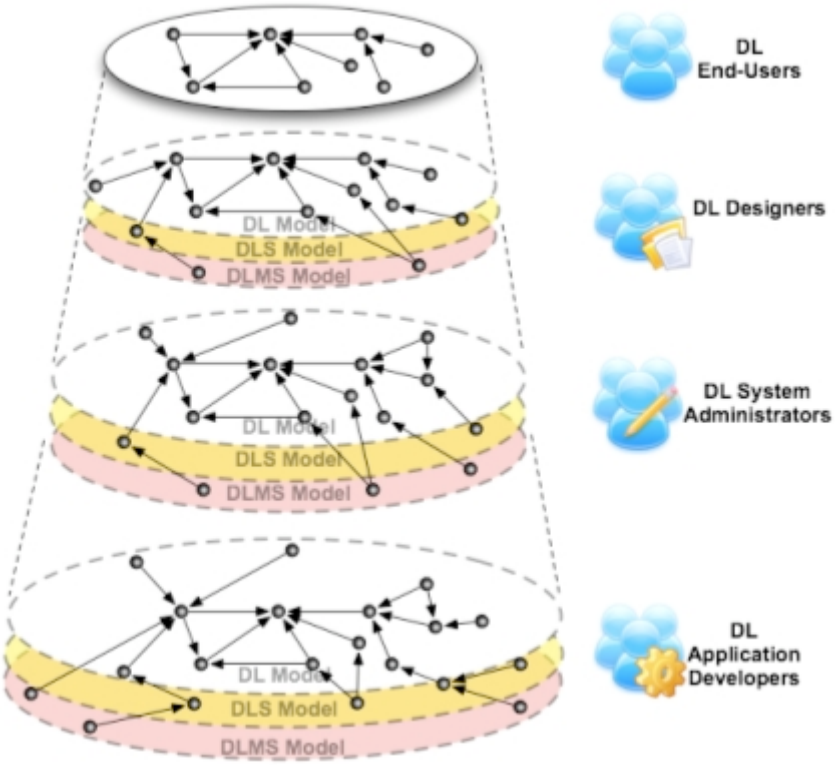

Figure 4. Hierarchy of Users' Views

5 Reference Frameworks

The digital library universe is complex and comprises multiple elements (see Figure 5). The representation of the details of these elements depends upon the introduction of frameworks supporting different levels of abstraction:

- **Reference Model** - As stated in [13], "A Reference Model consists of a minimal set of unifying concepts, axioms and relationships within a particular problem domain, and is independent of specific standards, technologies, implementations, or other concrete details". Digital libraries need to obtain a corresponding Reference Model in order to consolidate the diversity of existing approaches into a cohesive and consistent whole, to offer a mechanism for enabling the comparison of different DLs, to provide a common basis for communication within the DL community, and to help focus further advancement [17].
- **Reference Architecture** - The Reference Architecture is an architectural design pattern indicating an abstract solution to implementing the concepts and relationships identified in the Reference Model. There may be multiple Reference Architectures that indicate how to design DigitalLibraries Systems built on the Reference Model. For example, we might have one Reference Architecture for DLSs supporting DLs constructed by federating local resources and multiple organizations, and another one for personal DLs or for specialised applications.
- **Concrete Architecture** - At this level, the Reference Architecture is actualised by replacing the mechanisms envisaged in the Reference Architecture with concrete standards and specifications. For example, a Concrete Architecture may specify that the run-time environment deployed on the hosting nodes will be CORBA or the Web Services Application Framework, and that four specific communicating Web Services will implement the Search functional component.

The relationship of these three frameworks to the general digital library environment is shown

in Figure 5. At the top is the most abstract Reference Model, which guides the more specific Reference Architecture and Concrete Architecture further down. In turn, these should constrain the development and implementation of any actual system. The three reference frameworks are the outcome of an abstraction process that has taken into account the goals, requirements, motivations and, in general, the digital library market, as shown in the left-hand side of Figure 5, and the best practices and relevant research shown on the right-hand side of the same figure. When these frameworks are adopted and followed by the community, the resulting systems will be largely compatible with each other; the interoperability thus afforded will open up significant new horizons for the field.

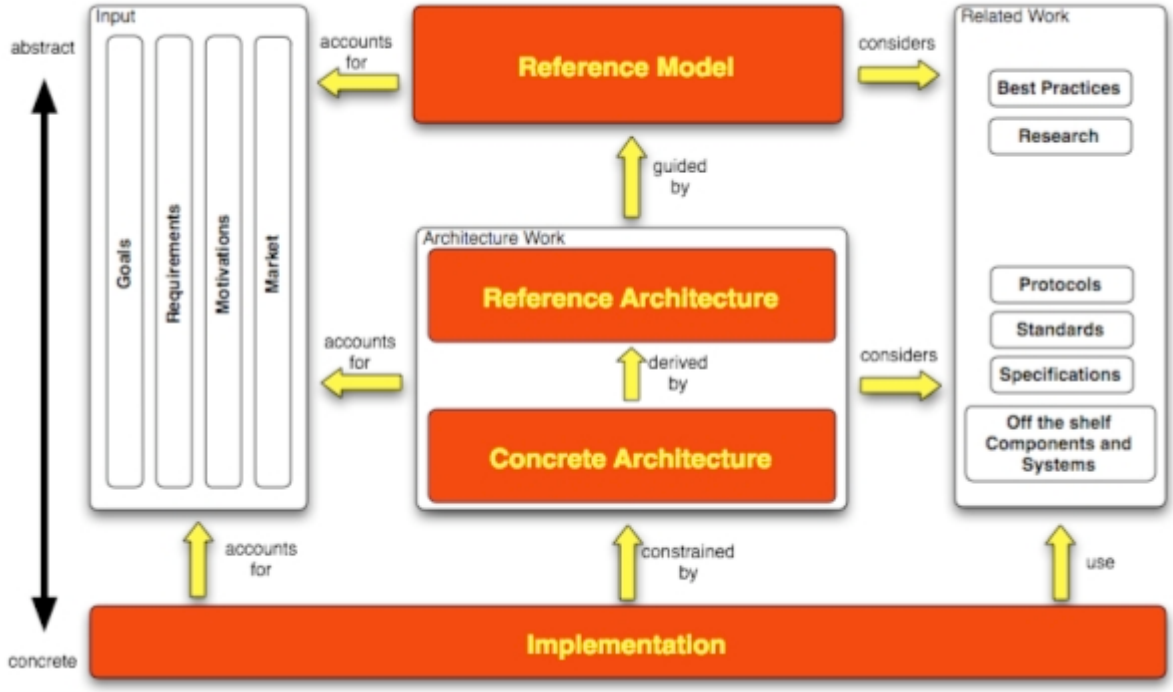

Figure 5. The Digital Library Universe[6]

6 Conclusions

This article has presented the core parts of *The Digital Library Manifesto* produced by members of the DELOS Network of Excellence on Digital Libraries to be used as a springboard for future foundational research and development in the domain of Digital Libraries.

The Digital Library Manifesto is accompanied by two other works currently available in their first release as a DELOS deliverable [1], introducing the DL Reference Model and a DLS Reference Architecture. We hope that these documents will provoke discussion and debate within the digital library community that will promote their improvement. Clearly, diversity of needs among different digital library "systems" will continue to introduce new concepts that will require incorporation into the Reference Model and in the Reference Architectures. Hence, these documents should be considered as first versions of otherwise dynamic documents that will keep evolving, having the *Manifesto* as a firm foundation.

Acknowledgements

The Digital Library Manifesto is built on the experience and knowledge gained by many previous efforts in the Digital Library area. In particular, the authors wish to acknowledge the considerable input to the drafting of the *Manifesto* received from the participants of the DELOS Reference Model Workshop held in Frascati, Italy, in June 2006. More details about this Workshop can be found at <http://www.delos.info/ReferenceModel>.

Note on website citations: All citations of websites were validated on 16 February 2007.

Notes

1. DELOS: Network of Excellence on Digital Libraries (G038-507618) is co-funded by the European Union under the 6th Framework Programme runs from January 2004 through December 2007.

2. The Third Manifesto, from the book 'Databases, Types, and the Relational Model: The Third Manifesto' by H. Darwen and C. J. Date, Addison-Wesley, 2007. <http://www.thethirdmanifesto.com/>.

3. To the extent that it is helpful, one may draw an approximate analogy between the world of Digital Libraries and the world of Databases. A DBMS (e.g., the DB2, Oracle system, MySQL or PostgreSQL) corresponds to a DLMS, offering general data management services. A DBMS together with all application software running on top of it at an installation corresponds to a DLS. Finally, a DL corresponds to a so-called "Information System" that consists of the above software, its data, and its users.

4. From here on, we shall use the terms "Digital Library" (or its acronym "DL"), "Digital Library System (DLS)" and "Digital Library Management System (DLMS)" to denote the systems identified in Sec.2, while by the term "digital libraries" we shall refer to the whole field of digital library research and applications.

5. This is an appropriate adaptation of the 'Architecture' definition from the Glossary of CMU's Software Engineering Institute. <http://www.sei.cmu.edu/opensystems/glossary.html>.

6. This picture was inspired by the "Reference Model for Service Oriented Architecture" document [13].

References

(All citations of websites were validated on 16 February 2007.)

[1] M. Agosti, L. Candela, D. Castelli, N. Ferro, Y. Ioannidis, G. Koutrika, C. Meghini, P. Pagano, S. Ross, H.-J. Schek, H. Schuldt. *A Reference Model for DLMSs* Interim Report. DELOS Deliverable n. 1.4.2, September 2006.

[2] Borgman C.L. What are digital libraries? Competing visions. *Information Processing and Management*, 35(3):227-243, 1999.

[3] BRICKS Integrated Project: Building Resources for Integrated Cultural Knowledge Services, <http://www.brickscommunity.org>.

[4] DELOS Network of Excellence on Digital Libraries, <http://www.delos.info>.

[5] Del Bimbo, A.; Gradmann, S.; Ioannidis, Y. (Eds.) *Future Research Directions*. 3rd DELOS Brainstorming Workshop Report, Corvara, Italy, July, 2004.

[6] DILIGENT Integrated Project: Digital Library Infrastructure on Grid Enabled Technology, <http://www.diligentproject.org>.

[7] Fox, E.A.; Akscyn, R.M.; Furuta, R.; Leggett, J.J. Digital Libraries. *Communications of the ACM*, 38(4):23- 28, April 1995.

[8] Fox, E.A.; Marchionini, G. Toward a Worldwide Digital Library. *Communications of the ACM*, 41(4):29-32, April 1998.

[9] Gonçalves, M.A.; Fox, E. 5SL - A Language for Declarative Specification and Generation of Digital Libraries. In *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'02)*, Portland, Oregon, USA, 2002.

[10] Ioannidis Y. (Ed.) *Digital Libraries: Future Directions for a European Research Programme*. DELOS Brainstorming Report, San Cassiano, Italy, June 2001.

[11] Ioannidis Y. Digital libraries at a crossroads. *International Journal of Digital Libraries*, 5(4):255-265, August 2005.

[12] Ioannidis, Y.; Maier, D.; Abiteboul, S.; Buneman, P.; Davidson, S.; Fox, E.; Halevy, A.; Knoblock, C.; Rabitti, F.; Schek, H.; Weikum, G. Digital library information-technology infrastructures. *International Journal of Digital Libraries*, 5(4):266-274, August 2005.

[13] MacKenzie, M.; Laskey, K.; McCabe, F.; Brown, P.; Metz, R. *Reference Model for Service Oriented Architecture*. OASIS Committee Draft 1.0, February 2006.

[14] Schek, H.-J.; Schuldt, H. DelosDLMS - Infrastructure for the Next Generation of Digital Library Management Systems. *ERCIM News*, No. 66, July 2006.

[15] Snowdon, D.N.; Churchill, E. F.; Frecon, E. (Eds.) *Inhabited Information Spaces - Living with your Data*. Springer, London 2004.

[16] Soergel, D. A Framework for Digital Library Research. *D-Lib Magazine*, 8(12), December 2002, <http://www.dlib.org/dlib/december02/soergel/12soergel.html>.

[17] The Digital Library Reference Model website, <http://www.delos.info/ReferenceModel>.

---

doi:10.1045/march2007-castelli