History, Evolution and Impact of Digital Libraries

CHAPTER · JANUARY 2011

Leonardo Candela
Italian National Research Council

Donatella Castelli
Italian National Research Council

Pasquale Pagano
Italian National Research Council

Chapter 1
History, Evolution, and Impact of Digital Libraries

Leonardo Candela
Consiglio Nazionale delle Ricerche, Italy

Donatella Castelli
Consiglio Nazionale delle Ricerche, Italy

Pasquale Pagano
Consiglio Nazionale delle Ricerche, Italy

ABSTRACT

Digital Libraries have achieved a fundamental role in our knowledge society. By making the wealth of material contained in libraries, museum, archives and any knowledge repository worldwide available they are giving citizens in every place of the world the opportunity to appreciate their global cultural heritage and use it for study, work or leisure. They are revolutionising the whole knowledge management lifecycle. In this chapter, the history characterizing these "knowledge enabling technologies" is described. The history starts from the early attempts toward systems supporting knowledge discovery and reaches the current age in which a plethora of different realizations of digital library systems coexist. The evolutionary process conducting to the current, multi-instanced and still evolving status of affairs as well as the motivations governing it are identified and presented. The main initiatives and milestones producing the nowadays instances of these knowledge enabling systems are mentioned. Finally, the impact these systems had and are having on various aspects of our society is discussed.

1. INTRODUCTION

Libraries, together with archives, have always been the primary institutions delegated to manage – collect, preserve and diffuse – human knowledge and culture. When advances in computer science allowed dealing with digital representation of documents dedicated to capture human knowledge and culture rather than printed ones, libraries were particularly involved in exploiting the potential of the digital revolution. Thus "digital libraries" soon became the term to indicate the digital counterpart of traditional libraries. However, digital library systems have greatly evolved since their early appearance. Today they have become complex networked systems able to support communica-

DOI: 10.4018/978-1-60960-031-0.ch001

tion and collaboration among different worldwide distributed communities, dealing with "digital objects" comprising not only the digital counterpart of printed documents, but also images, video, programs and any other kind of multimedia objects a community may define as appropriate to its working and communication needs.

The evolution of digital libraries (DLs) has not been linear, coming from the contribution of many disciplines. This has created several conceptions of what a DL is, each one influenced by the perspective of the primary discipline of the conceiver(s) or by the concrete needs it was designed to satisfy. As a natural consequence, the "history" of Digital Libraries, which is now approximately twenty years long, is the history of a variety of different types of information systems that have been called "digital libraries". These systems are very heterogeneous in scope and functionality and their evolution does not follow a single path. In particular, when changes happened this has not only meant that a better quality system was been conceived superseding the "preceding" ones but also meant that a new conception of digital libraries was born corresponding to new raised needs. As it will be seen, most of the systems dealt with in this history are still living in their original conception, even though not in their original technological solutions.

The rest of this chapter goes back over this history, giving an account of past and present understanding of these kind of systems and on-going work in the area. The chapter concludes with a vision of the impact that new DLs are expected to have in the near future.

2. DIGITAL LIBRARIES: THE EARLY TIMES

The digital library concept can be traced back to the famous papers of foreseer scientists like Vannevar Bush and J.C.R. Licklider identifying and pursuing the goal of innovative technologies and approaches toward knowledge sharing as fundamental instruments for progress. Bush (Bush, 1945) devised "*a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility.*". Moreover, on top of it there is "*a transparent platen. On this are placed longhand notes, photographs, memoranda, all sorts of things*". Because of the lack of digital support, he identified in "*improved microfilm*" the means for content storage and exchange: "*contents are purchased on microfilm ready for insertion. Books of all sorts, pictures, current periodicals, newspapers, are thus obtained and dropped into place*". Of course, he envisaged also support for knowledge discovery ("*provision for consultation of the record by the usual scheme of indexing*"), access ("*to consult a certain book, he taps its code on the keyboard, and the title page of the book promptly appears before him*") and management ("*new forms of encyclopedias will appear, ready made with a mesh of associative trails running through them, ready to be dropped into the memex and there amplified*"). Licklider realized that computers were getting to be powerful enough to support the type of automated library systems that Bush had described and in 1965, wrote his book (Licklider, 1965) about how a computer could provide an automated library with simultaneous remote use by many different people through access to a common database. Because of this, Licklider is also considered a pioneer of Internet and in its book he established the connection between Internet and digital library. Thus, it is not surprising that research and development activity on digital libraries started in the early 1990s, with the Internet proliferation, and that Internet has created unprecedented possibilities to discover and deliver human knowledge.

The first systems delivering knowledge artefacts in digital form can essentially be seen as archives of digital texts accessible through a search service and implemented by a centralized metadata catalogue.

An important example of a system conceived to respond to concrete needs went on-line starting in August, 1991. This system, originally named "e-print archive" and now worldwide known as *arXiv*, was born as an experimental means for making scientific communication more effective and economic, a requirement mostly expressed by the physicists community.(Ginsparg, 1994) Although tight to the technologies of those years, this system provided a paradigm for changes in worldwide, discipline-wide scientific information exchange, even though its rapid acceptance was facilitated by the pre-existing "preprint culture" of the community of high energy theoretical physics, in which the irrelevance of refereed journals to ongoing research has long been recognized.

The arXiv system opened the way to deal with the social and economical issues related to the open access to outputs coming from publicly funded research, that were later officially stated in the Berlin Declaration (http://oa.mpg.ed, (n.d.)), in 2003, and now strongly promoted by many initiatives and Funding Agencies (Council of the European Union, 2007). It can be considered as the prototype of (institutional) repository systems (Lynch, 2003), i.e. systems characterized by functionality for managing self-publishing (document submission, reviewing, editing, etc.) and dissemination of born digital documents.

The early ones of such systems were constructed on a rather simple architecture, with the exception of very few cases. This worked to the advantage of their diffusion and adoption by different scientific communities. Besides arXiv, significant examples of such early systems were archives of various type like Electronic Thesis and Dissertations repositories (*ETDs*), whose pilot project started in 1996 (Fox, Eaton, McMillan, et al, 1996), and archives of cognitive sciences papers (*CogPrints, (n.d.)*) and of research papers in economics (*RePEc, (n.d.)*) both launched in 1997. The former was a system which was offering services for submitting, browsing and searching electronic thesis in PDF format. The availability of this product stimulated the creation of the Networked Digital Library of Theses and Dissertations (*NDLTD, (n.d.)*) international organization, still operational, which registers and keep track of ETDs.

CogPrints, was initially conceived as repository allowing the cognitive science community to self-archive their papers. It now contains more than 3,000 artefacts starting from 1950. In 2000 it was made compliant with the protocol defined by the Open Archives Initiative (see Section 3) and then its software was converted into the EPrints Digital Repository Software (EPrints, (n.d.)), a flexible platform supporting easy and fast set up of repositories of open access research outputs. Because of its simplicity, EPrints is currently widely used, more than 250 repositories declared to rely on it.

Similarly, RePEc was initially conceived as an open repository of electronic papers in a specific domain. Thomas Krichel, principal investigator of the RePEc Project, in 1997 illustrated the principles underlying a new realised version of this system by affirming *"Distributed archives should offer metadata about digital objects (mainly working papers); the data from all archives should form one single logical database despite the fact that it should be held on different servers; users could access the data through many interfaces; providers of archives should offer their data to all interfaces at the same time."* Krichel, with these statements was anticipating a view that would have largely emerged few years later.

These systems – all still living in more recent and enhanced versions – represent very embryonic forms of digital libraries. In fact, their functionality is essentially confined to (self-)publishing of simple information objects and discovery of these information objects through rudimentary search and browse facilities.

In parallel with the repository systems, other kinds of systems, sharing with them the need for supporting digital documents storage and retrieval but oriented to enlarge the pool of services and functionality offered to their clientele, started being designed and developed. The founding ideas for these systems were extending and enhancing the capabilities of information and storage systems so that they could manipulate and deliver rich digital artefacts besides documents' bibliographic description, *i.e.* metadata. Essentially, new system development activities started with the goal of supporting scholars by providing them with the functionality of a traditional library (collect, store, organise and discovery information) in the context of distributed and networked collections of digital information objects in user-friendly ways (Belkin, 1999). The initiatives that started giving live to such systems, that can be reasonably considered as substantial digital libraries, were the Digital Library Initiative (DLI) in the US, while national initiatives, *e.g.* eLib in UK, and EU funded projects including a dedicated Network of Excellence, DELOS (DELOS, (n.d.)), have characterised the European scene (Griffin, Peters, and Thanos (2005).

The Digital Library Initiative (DLI) consisted of two major competitive funding programs, the first of which started in 1994 and funded six research projects (chosen among 73 proposals) over a four-year period (Schatz and Chen,1996) while the second phase was dedicated to extend the research carried out during the previous phase by including content providers thus to guarantee the availability of real testbed to validate research outcomes. However, the DLI funded projects have not been the only ongoing efforts (CACM, 1995) even if they were very innovative because they focused on future technological problems. The six projects funded by DLI phase one were: the *California Environmental Digital Library* (Wilensky, 1995) focused on developing the technologies to access large, distributed collections of photographs, satellite images, videos, maps, documents, and "multivalent documents" and to support work-centred digital information services (Wilensky, 1996); the *Alexandria Digital Library* (Smith and Frew, 1995) focused on building an online, distributed digital library for geo-referenced[1] information, including maps, aerial photographs, satellite imagery, and catalogue records, and on supporting geographically defined queries (Smith, 1996); the *Informedia Digital Video Library* (Christel, Kanade, Maudlin, et al., 1995) focused on establishing a large, online digital video collection with full-content and knowledge-based search and retrieval (Wactlar, Kanade, Smith and Stevens, 1996); the *Interspace* (Schatz, 1995) focused on building a large collection of technical engineering and physics literature that can be searched effectively across multiple indexes with a single interface (Schatz, Mischo, Cole, et al., 1996); the *University of Michigan Digital Library* (Crum, 1995) focused on creating a digital library architecture based on the notion of software "agents" (Atkins, Birmingham, Durfee, et al., 1996); the *Stanford Digital Library Project* (Stanford Digital Libraries Group, 1995) focused on addressing aspects of interoperability over heterogeneous services and collections via the "InfoBus" protocol, which provides a uniform way to access a variety of services and information sources through "translators" (Paepcke, Cousins, et al., 1996).

Despite none of these systems exist anymore as a running service[2], the solutions proposed, the technology developed as well as the resources collected and built have been largely used by more complex DLs developed later. It is well known that one of the most important success stories resulting from these projects is *Google*®. Page and Brin started working on their search engine while being PhD Students at Stanford working on the Stanford Digital Library Project. Actually, the Digital Library Initiative merits goes far beyond the specific work that it funded and we can affirm that it gave shape to "digital library" as a new research discipline. Research in digital library topics was not new but it had been fragmented across many disciplines. This program led to conferences, publications and researcher teams explicitly interested in doing research in digital libraries. Moreover, it gave directions to the overall movement toward a practical research field.(Arms, 2001)

As anticipated, in Europe the scene was characterised by the existence of *DELOS* initiatives. The activities of DELOS started with the "DELOS Working Group" at the end of the 1990s[3], and the DELOS Thematic Network, under the Fifth Framework Program (2000 – 2003). Since its beginning, the main objective of DELOS was to advance the state of the art in the field of digital libraries by coordinating the effort of the major European research teams conducting activities in the main fields of interest. One of the early important achievements was the establishment of a formal collaboration with the US National Science Foundation and the creation of five joint EU-US collaborative Working Groups. These working groups explored DL-related technical, social and economic issues, and published a set of recommendations with respect to DL interoperability, metadata, IPRs and economics, global resource discovery and multilingual information access in a special issue of the International Journal of Digital Libraries (Griffin, Peters, Thanos, 2005). The last phase of the DELOS evolution was its transformation into the DELOS Network of Excellence, under FP6 (2004 – 2007). Its mission was to integrate and coordinate the on-going research activities of the major European research teams in the field of Digital Libraries. The main achievement was the definition of the "DELOS DL Reference Model" (Candela, Castelli, et al., 2007), a formal and conceptual framework describing the characteristics of the Digital Library domain. The main merits of DELOS are represented by its significant contributions to the creation of a European DL research community, by the organisation of important durable scientific events and infrastructures (*e.g.* ECDL[4], CLEF[5], INEX[6]), and by notable suggestions, in the form of either prototypes or roadmap reports, anticipating many actions of the European Commission in the field of Digital Libraries (Thanos, 2009).

In parallel with the DELOS initiatives, in Europe activities dedicated to the development of exploratory systems going in the direction of "true" digital libraries started with the support of the European Commission programmes[7]. Among the projects initially funded, notable are those described in the following. The *European Chronicles On-Line (ECHO)* (Savino and Peters, 2004) focusing on the development of a digital library service for historical films by using an open architecture approach distributing digital film archive services. In addition, it was intended to develop new models for intelligent audio-visual content-based searching and film-sequence retrieval, new video abstracting tools, and user interfaces specifically tailored to the new functionality. The provision of multilingual services and cross language retrieval tools was also addressed. Another project, i.e. *An Integrated Art Analysis and Navigation Environment (ARTISTE)* (Allen, Vaccari and Presutti, 2000), focused on giving providers, publishers, distributors, rights protectors and end users of art images information, as well as the multi-media information market as a whole, a more efficient system for storing, classifying, linking, matching and retrieving art images. This environment was providing, for example, automatic extraction of metadata based on iconography, painting style, etc; content-based navigation for art documents; distributed linking and searching across multiple archives allowing ownership of data to be retained; and storage of art images using large multimedia object relational databases. The *Collaboratory for Annotation, Indexing and Retrieval of Digitized Historical Archive Material (COLLATE)* (Thiel, Brocks, Frommholz, et al., 2004) project focused on the development of a collaborative work environment for archives, researchers and end-users focused on historic film documentation, including censorship files, photos and film fragments in which users take an active part in evaluating sources and adding valuable information.

Being dedicated to build exploratory systems, both the DLI funded projects and the FP5 funded projects spent the majority of their effort in implementing proof-of-concept systems by integrating results from various and separate research fields and experimenting these solutions in a specific context. Thus, each project was dedicated either to serve the need of a specific community or to design and implement a certain functionality over a specific kind of information. Not surprisingly, the majority of first-generation digital library systems were "from scratch", "monolithic applications"[8] lacking of reusability, ease of installation, customisation and configuration. (Ioannidis, Y., Maier, D., et al., 2005)

Among the first attempts to overcome the monolithic approach notable are *NCSTRL* (Davis, J,R, Lagoze, C., 2000), the Networked Computer Science Technical Research Library, and its enabling technology *Dienst* (Davis, J.R., Lagoze, C., 1995). Dienst was based on quite innovative principles at the time in the digital library domain, namely: open architecture, federation and distribution. According to these principles: the functionality of a digital library system were available in the form of distinct functional units, each exposing its operational semantics through an open protocol; digital library systems are compositions of these functional units and new functionality can be added through the implementation of value-added services, which interact with existing others using established protocols; the components (and content) of a digital library could be spread over the global Internet, but should be presented to the user as a single system. NCSTRL grew a lot in the United States. Approximately three years after its inception, the NCSTRL collection contained about 22,000 documents from 118 different institutions.

In August of 1995, ERCIM, the European Research Consortium for Informatics and Mathematics, asked to join the NCSTRL network. This gave birth to *ETRDL*, the *European Technical Report Digital Library* (Biagioni, S., Borbinha, et al., 1998). This "expansion" of NCSTRL raised reliability and performance problems due to connectivity characteristics of the global Internet. To overcome these issues and obtain good performance, the Dienst initial architecture was modified by adding the notions of *collection service* and *connectivity region* (Lagoze, C., Fielding,. D, 1998). ETRDL was also the first important experience in Europe in designing and operating a digital library having a European scale. In collecting requirements from the ERCIM community it became evident that this community had its own specific requirements (Andreoni, A., Baldacci, M.B., et al., 1999), not all of which were covered by the basic Dienst system as adopted by NCSTRL. The list of requirements included three important aspects: the need for classification mechanisms; the need to cater for languages other than English and the need to provide on-line document submission facilities. The ETRDL supporting technology was designed and implemented by maintaining interoperability with NCSTRL, so that users could perform cross-Atlantic searches, while at the same time extending this system to provide additional functionalities as requested by ERCIM users. Among the new functionalities, on-line document submission distinguished ERCIM from most of the contemporary DL systems. These were conceived to serve end-users only as consumers of information, and submission was usually performed outside the DL by means of specific procedures operated by either the author or a librarian. ETRDL engaged digital library designers in a lot of relevant choices. Most were technical ones, but some related to policy and administration. Most of the large European Initiatives funded few years later were the result of this early experience.

The projects and initiatives described so far characterised the early times of the digital library domain, the birth of the field. Once established, the field evolved like any other research and development field. The evolution has been multi-faceted and spontaneous, thus leading to the today status in which, despite the existence of a reference model (Candela, L., Castelli, D., et al., 2007), the term "digital library" continues to evoke different impression in each digital library practitioner exactly like in the past (Fox, E.A, Akscyn, R.M., et al., 1995). In the rest of the chapter the evolution of the field is described by clustering the main initiatives and projects in three main categories: those having large-scale content sharing as guiding principle, those dedicated to the definition and development of generic software systems for simplifying the building and operation of digital libraries (*Digital Library Management Systems*)(Ioannidis, 2005) and those leading to new research environments in which all researchers have shared access to scientific facilities including data, instruments, computing and communications regardless of their location in the world (a.k.a. *e-Infrastructures*)[9].

3. DIGITAL LIBRARIES EVOLUTION: CONTENT SHARING

The construction of digital libraries similar to those just described was very resource-consuming since, for each new one, both the content and the software providing its functionality were built from scratch. At the end of the 1990s, the experiences of using distributed architectures to implement proper digital libraries and the proliferation of independent repositories of valuable content stimulated the idea of reusing content already collected (and curated) in existing independent repositories so as to reduce the effort to build large-scale digital libraries. However, many obstacles were to be solved to fully implement this solution. The major of them was certainly how to implement repository service interoperability, *i.e.* the capability of seamlessly accessing and using the content managed in distributed and heterogeneous repositories.

Approaches based on cross-searching multiple archives based on a common protocol, such as *Z39.50*[10], (Miller, P., 1999) were considered at the time costly and hardly scalable. A very important meeting toward the interoperability of electronic repositories was organised in Santa Fe, New Mexico, on October 1999, with the goal to establish recommendations and mechanisms to facilitate cross-archive value-added services. This meeting led to the *Santa Fe Convention* – a combination of organizational principles and technical specifications to facilitate a minimal but potentially highly functional level of interoperability among scholarly e-print archives – and to the establishment of the *Open Archives Initiative*. (Van de Sompel, H., Lagoze, C., 2000) The meeting started by discussing a concrete example of interoperability implemented through the UPS Prototype (Van de Sompel, H., Krichel, T., Nelson, M.L., 2000) and recognising its potentialities. The UPS prototype demonstrated the integrated action of a variety of services operating over data originating from a set of archives. Each of those services provided a reasonably rich level of functionality (accessible through a set of protocol methods). The participants recognised that trying to reach consensus on the full functionality of the prototype was "aiming too high" and that a proper degree of modesty in the approach toward integration capable to balance the cost of participation with the need for adequate functionality was mandatory. The Santa Fe Convention identified two key roles in participating institutions: "data providers" and "service providers". Data providers were in charge to handle the depositing and publishing of resources in a repository and "expose" for harvesting the metadata (what they called *record*) about resources in the repository. They were the creators and keepers of the metadata and repositories of resources. Service providers were in charge of harvesting metadata from data providers for the purpose of providing one or more services over the collected data. The types of services that might be offered included a search interface, peer-review system, etc. The cooperation between content and service providers was regulated by a protocol, initially defined as a subset of the Dienst protocol and nowadays known as the *Open Archive Protocol for Metadata Harvesting (OAI-PMH) (*Lagoze, C., Van de Sompel, H., 2001*)*. This is a simple protocol made by six protocol requests and responses and because of its simplicity and relatively low cost of adoption it is so diffuse as to become a sort of *de-facto* standard solution.

One of the first experiments of implementing a large-scale digital library search service across multiple data providers was performed by *TEL*, The European Library project, which started in 2001 (Woldering, B., 2004). The key aim of TEL was to investigate the feasibility of establishing a new pan-European service which would ultimately give access to the combined resources of the national libraries of Europe. The technical issue at the beginning of the project was the heterogeneous nature of access to the data of the partner libraries: some offered access to bibliographic data via the Z39.50 protocol, some did not. Furthermore, not all collections were included in the Online Public Access Catalogues (OPACs) (Altelman, K., Lineman, E., Pace, A.K., 2006) of the national libraries. The first task for TEL was to find a solution for pooling the metadata of all collections and for offering for integrated search. A solution was firstly identified in using the Z39.50 protocol for OPACs and the Hypertext Transfer Protocol for the metadata not offered via Z39.50. After the publishing of OAI-PMH, this protocol was adopted by TEL for the harvesting of metadata for the central index of those resources not available via Z39.50. TEL was finished in 2004 and now delivers a web service for accessing the combined resources (books, magazines, journals, etc. – both digital and non-digital) of the forty-five national libraries of Europe. It offers free searching and delivers digital objects – some free, some priced.

Another important initiative for large-scale cross-repository services was *DARE*, the Digital Academic REpositories (Kuil van der, A. and Feijen, M., 2004). Started in 2003, this was a joint initiative by Dutch Universities, National Library of Nederland, and other Dutch Organizations. Its aim was to store the digital outcome of all Dutch research in a common network of Institutional Repositories (IRs) (Lynch, C.A., 2003) in order to facilitate its dissemination. DARE went towards the construction of a federation of IRs by providing a set of guidelines for the cooperation and interoperability of otherwise independent IRs. The guidelines imposed a set of standards at the data level to which the participating repositories must line-up so as to enforce interoperability and enable the realization of services operating over the federation. Basically, DARE referred to OAI-PMH and adopted simple Dublin Core[11] as the mandatory metadata set, plus DARE-qualified Dublin Core as an optional metadata set. IRs should convert their internal metadata format to the DARE metadata format and provide an expose their records through the OAI-PMH protocol. No particular document format or model was imposed to the repositories, but digital objects should be reachable for harvesting via HTTP links or through a jump-off page. Since June 2008, the DARE service can be accessed through the NARCIS portal[12].

In the US, the National Science Foundation funded the *National Science Digital Library (NSDL)* (Zia, L.L., 2001) with the aim to provide organized access to high quality resources and tools that support innovations in teaching and learning at all levels of science, technology, engineering, and mathematics education.[13]

These large-scale initiatives devoted to aggregate in a single place knowledge that is spread across a plethora of archives and systems will ever exist for a series of reasons including the existence of various (institutional) repositories and the ever growing multidisciplinary nature of our society. In particular, TEL and DARE anticipated important initiatives, namely, Europeana and DRIVER, respectively, which were launched few years later.

Europeana[14] is a Thematic Network funded by the European Commission under the eContentplus programme, as a part of the i2010 initiative[15]. Europeana began in July 2007. Originally known as the European digital library network – EDLnet – it is the result of a partnership of 100 representatives of heritage and knowledge organisations and IT experts from throughout Europe. Objective of Europeana is to provide access to Europe's cultural and scientific heritage through a cross-domain portal. The first Europeana prototype, launched in November 2008, provided simple search and retrieval facility on an information space of approximately two millions of digital objects selected from Europe's museums, libraries, archives and audio-visual collections, harvested through the OAI-PMH protocol. The first production quality version of Europeana (called Rhine) will go live on July 2010, to be followed in April 2011 by a more sophisticate version (Danube), including more contents and offering a richer set of functionality. The intention is that by 2010 the Europeana portal will give everybody direct access to well over 6 million digital sounds, pictures, books, archival records and films. Moreover, Europeana's goal is to realize a system serving very different type of users. It should meet occasional curiosity of generic users as well as the information needs of school children and students. It should also provide academic students and teachers with certified information and the possibility to export information for courses, as well as offer expert researchers and professional the possibility of searching, verifying and annotating information and using ad-hoc services. In the context established by Europeana, special type of providers are the *aggregators*, *i.e.* specialised DLs that act as collectors of content from other providers. For instance, Culture.fr is the largest aggregator, providing content from about 480 organizations in France, including the Louvre and the Musée d'Orsay. The information resources that populate Europeana's information space are harvested as *surrogates* of the original objects that are located at content providers' sites. Since surrogates may also contain elements of the original object (table of contents, full text index items, music and video abstraction etc.), the very interesting new feature of Europeana is that it will also deliver digital objects besides metadata. Clearly, heterogeneity and interoperability are main issues that such a DL is having to deal with, as well as, of course, with scalability, quality of service and, more in general, sustainability of the joint portal.

DRIVER[16] is another notable example of a DL that relies on content provided by a large number of external data providers. It is the result of two subsequent projects funded by the European Commission in the period 2006-2009. The main aim of these two projects is to create the organisational and technological conditions for the set up of a European Repository Infrastructure (Jones, S., Manghi, P., 2009). The main instrument identified by the project to address organisational issues is the DRIVER Confederation[17]. The Confederation partners represent European and international repository communities, like subject based communities, repository system providers, service providers, as well as political, research, and funding organisations, who share the DRIVER vision to allow all research institutions in Europe and worldwide to make all their research publications openly accessible through institutional repositories. In the spirit of this shared goal, the DRIVER confederation encourages a combined effort of repository development by setting up guidelines and best practices that favour the realization of a shared, trusted, long-term repository infrastructure. From the technical point of view, DRIVER is based on the *D-Net* technology[18]. This enabling technology is quite innovative in the context of these kinds of aggregative systems because it is oriented to the realisation of a digital library infrastructure (cf. Sec. 5). D-Net is based on a Service-oriented architecture, where distributed and shared resources are implemented as standard Web Services and applications consist of sets of interacting services. It offers services to both data providers, that through it can more easily share their content, and service providers, that are facilitated in implementing DLs that exploit the aggregated content.[19] At the time of this writing, the DRIVER service provides access to approximately one million records out of 200+ repositories across 27 countries. Moreover, it delivers three DL applications: the Belgium national repository portal, offering search over the Belgium Repository Federation subset; Recolecta national repository portal, offering search on the Spanish Repository Federation subset; and the main DRIVER portal, providing access and advanced functionality over the whole space.

The current Europeana and DRIVER services operate an information space of metadata records, i.e. they harvest metadata records through the OAI-PMH protocol from exiting repositories and then they run their services by exploiting this content. Because of this they suffer from the limitations that OAI-PMH poses if it has to be used to exchange information objects that are "rich" in structure and payload as those at the core of changing nature of scholarship and scholarly communication.(Van de Sompel, H., Payette, S., 2004)(Van de Sompel, H., Lagoze, C., 2006) In particular, when feasible, they give access to the content associated with the metadata by exploiting URL or some other information contained in the record. This solution to access information objects, however, suffers of two main problems: (*i*) the access is not always feasible since there is no standard protocol to access objects; (*ii*) there is no way of accessing compound objects since the structure and the relations holding among the different parts is unknown. A solution to this problem may come from the *OAI-ORE*[20] standard, whose version 1.0 has been released in October 2008 by the Open Archives Initiative. This standard, based on Web standards, proposes a solution to handle aggregations of Web resources. These aggregations, sometimes called compound digital objects, may combine distributed resources having multiple media types including text, images, data, and video as to form innovative research outcomes.

Both Europeana and DRIVER have already planned to move very soon to technologies *à la* OAI-ORE to manage compound objects.

All the systems and initiatives described in this section are essentially oriented to *content sharing*. Moreover, the majority of them is characterised by a strong organisational effort since the model is based on a cooperative participation of the content providers. Content sharing across digital libraries is now being largely promoted as an important strategy to reduce the digital library set up costs largely coming from selecting, digitising, describing, and digitally curating content resources. However, the realisation of wide and generalised content sharing is today still problematic due to the great variety of proprietary models and ontologies adopted by existing systems and by the lack of systematic approach to interoperability. DL.org (Castelli, D., Parker, S., 2009), a recently funded EC project stemming from the DELOS project, is paving the way for the future interoperability of DL systems thus making feasible the implementation of global digital library infrastructures.

4. DIGITAL LIBRARY EVOLUTION: DIGITAL LIBRARY MANAGEMENT SYSTEMS

The reuse of content is not the only strategy that has been put in place in order to reduce the cost of DL development. Another important step toward this aim has been the conception of digital library management systems (DLMSs), *i.e.* systems that provide the appropriate framework to both (i) produce and administer a Digital Library System incorporating the suite of functionality considered fundamental for Digital Libraries and (ii) integrate additional software offering more refined, specialised or advanced functionality.(Candela, L., Castelli, D., et al., 2007) Thus a DL can be built by configuring and deploying a DLMS and then loading or harvesting content. This approach largely simplifies and reduces the effort required to set up a DL and, generally, guarantees a better quality of service.

These generic systems have started to appear from the beginning of 2000 even though implementing the devised DLMS features only to some extent. The major characteristics that distinguish them from each other are the class of functionality offered, the type of information object model supported, and the openness of their architecture.

Repository management systems, i.e. systems specifically dedicated to implement repositories, represent a first primitive form of DLMSs. Usually, these systems are configurable to an extent that varies a lot from system to system, offer limited functionality to the administrators for managing the system once it has been installed; further, they are centralised and rarely extensible.

One of the first exemplars of these software systems is *Greenstone* (Witten, I.H., Bainbridge, D., 2002). This system, was conceived for simplifying the construction and presentation of information collections by offering standard search and browse facilities. Its simplicity, easy to use and the activity conducted by the conceivers to promote the usage of digital libraries in developing countries as a mean to actively participate in the information society have been very important factors toward the dissemination of this software.

Another main representative of this class of systems is *DSpace* (Smith, M., Barton, M., et al., 2003). This system, jointly developed by the MIT Libraries and Hewlett-Packard Labs starting from 2000, was conceived as an open source digital repository software for research institutions. The objective of its designers was to realize a system that could: (*i*) enable organizations to capture and describe digital material using a submission workflow module, or a variety of programmatic ingest options; (*ii*) support the distribution of an organization's digital assets over the web through a search and retrieval system; and (*iii*) preserve digital assets over the long term. (Tansley, R., Bass, M., Smith, M., 2003) The organisation of the information space in DSpace is intended to reflect the structure of a typical research organization. Each DSpace repository is organised in communities, each corresponding to a laboratory, research center or department. Communities contain collections, which are groupings of related content, and each collection is composed of items, which are the basic elements of the repository. Many instances of DSpace are currently operational, widespread all over the world.[21] DSpace is specifically appreciated for its effectiveness and for the simplicity of its installation and operation procedures. However, it is suitable only for very specific application domains due to its limited flexibility.

Flexibility has been, instead, the major goal addressed in designing the *Fedora* (*Flexible Extensible Digital Object Repository Architecture*) (Payette, S., Lagoze, C., 1998) system, more or less in the same period in which DSpace was introduced. Fedora was originally designed by the Digital Library Research Group at Cornell University under a NSF Grant and then its development proceeded as a collaboration with the University of Virginia Library funded by the Andrew W. Mellon Foundation. Differently from other repository systems designed as turn-key, vertical applications for storing and manipulating information objects through a fixed user interface, Fedora has been conceived to act as the foundational layer for a variety of multi-tiered systems, service-oriented architectures, and end-user applications. This means that Fedora was conceived as a service to be used programmatically for building more sophisticated applications. To meet this goal it has been implemented as a set of web services that provide full programmatic management of information objects as well and search and access to multiple representations of them. (Payette, S., Thorton, S., 2002) Also, the Fedora information object model is extremely flexible. It supports the expression of many kinds of compound objects. Objects are units of content which can include digital resources, metadata about the resources, and linkages to software tools and services (*disseminators*) configured to deliver the content in desired ways, even by producing it dynamically. (Lagoze, C., Payette, S., Shin, E., Wilper, C., 2006) In the original plans of its designers, the Fedora system was intended as a first element of a more complex service framework. This framework should consist of a set of loosely coupled services that interact and collaborate with each other. These services are expected to provide additional functionality that is not considered a fundamental function of a repository. Typical examples are the Fedora OAI provider and the Fedora Search service. Outside of the boundaries of the Fedora framework there are external services that can either call upon Fedora services or be leveraged by Fedora services in some way. The distinction between services within the Fedora Service Framework and those outside consists in that those within the framework are in a trusted relationship with the Fedora repository service and are designed to specifically interact with Fedora repositories, while services outside the framework are typically general-purpose services, or organization-specific services that call upon Fedora as an underlying repository for digital content.

Very recently (May 2009), the providers of DSpace and Fedora decided to create a new organisation, *DuraSpace*[22], with the goal to yield leadership and innovation in open source technologies for global communities who manage, preserve, and provide access to digital content. DuraSpace will sustain and grow its flagship repository platforms – Fedora and DSpace – and will also expand its portfolio by offering new technologies and services that respond to the dynamic environment of the Web and to new requirements from existing and future users. DuraSpace will focus on supporting existing communities and will also engage a larger and more diverse group of stakeholders in support of its not-for-profit mission.

Both DSpace and Fedora have essentially been conceived as management systems delivering digital repository functionality. Among the first management systems departing from this notion of exclusively repository-oriented system is *OpenDLib*,(Castelli, D., Pagano, P., 2002) developed at the Italian National Research Council, in Pisa. The design and development of OpenDLib was initiated in 2000 as a response to a pressing request for a general purpose software that could be customized to meet the needs of different DL application scenarios. It was explicitly designed to: (*i*) provide basic DL services to support the submission, description, indexing, searching, browsing, retrieval, access, preservation, and visualization of information objects; (*ii*) offer other digital library specific services, such as the ones providing the enforcement of access policies on information objects and the management of "user-shelves" able to maintain information objects versions, result-sets, session results, and other information; (*iii*) support plug-and-play expansion, thus making the systems capable of growing over time along several dimensions, not only along services, but also metadata formats supported, hosting servers, user communities addressed, and so forth.(Castelli, D., Pagano, P., 2003) Moreover, OpenDLib supports a powerful and flexible information object model (Candela, L., Castelli, D., et al., 2003), capable of representing structured, multilingual, and multimedia objects in a way that can be customized according to which content has to be handled. Further, it has introduced the notion of *virtual collections*,(Candela, L., Castelli, D., Pagano, P., 2003) *i.e.* collections not necessarily corresponding to an existing physical one, each characterized by its own access policy and dynamically update with new content whenever new objects matching the collection's membership criteria become available. In addition to the set of functions dedicated to serve the DL end-user that publishes and seeks for information, OpenDLib also provides a number of functions to support DL administrators in preserving objects, in applying object reviewing processes, in handling users and user group profiles and in deploying and managing services hosted by distributed servers. The introduction of these functions is actually the novelty of OpenDLib that makes it the first real exemplar of the class of software that the DL research community started later to name Digital Library Management System (Ioannidis, 2005) (Candela, L., Castelli, D., et al., 2007), slightly changing the attribute of "Digital Library Service System" that originally characterized OpenDLib.

Other Digital Library Management Systems departing from the notion of repository-oriented ones started to be developed since the 2000. In fact, as soon as the digital library development and use were proceeding, it became evident that the digital context was offering innovative possibilities that were not conceivable in the library world. Indeed, digital libraries could also become a major vehicle to support the entire cycle of scientific production, which comprise not only retrieval of relevant information, but also the analysis of this information and the production of new content that is then published and disseminated for use to others. Early systems that implemented this vision were developed in the framework of the *Scholnet* and *Cyclades* projects (Castelli, D., Pagano, P., Straccia, U., 2001), both funded by the EU 5th Framework Programme at the beginning of 2000. Both projects were aimed at extending the role of a digital library by providing services to support remote communication and collaboration among scholars. *Scholnet* was conceived to implement an enhanced set of specialised services enabling the immediate dissemination and accessibility of technical documentation within a globally distributed multilingual community. Accordingly, Scholnet was provided with the capability of delivering traditional services on multimedia documents such as videos of tutorials or seminars (possibly synchronized with corresponding textual slides), but also with innovative services such as handling document annotations. This service allowed different people to annotate documents with textual notes, ratings, links, etc. associated with either the entire document or with its parts, making annotations accessible publicly or restricted to groups. In addition, Scholnet provided a cross-language search facility permitting users to query in their own language and retrieve documents in other languages, as well as an automatic personalised information dissemination service, sending messages to the users potentially interested in newly arrived documents. *Cyclades*, instead, was conceived to realise an open collaborative virtual archive service environment supporting both single scholars as well as scholarly communities in carrying out their work. (Straccia, U., Thanos, C., 2004) In particular, it provided functionality to access large, heterogeneous, multidisciplinary archives compliant with the OAI-PMH standard (Lagoze, C., Van de Sompel, H., 2001) and distributed over the Web. Distinguishing functionality regarded collection mechanisms (for dynamically structuring the overall information space into meaningful, from some community's perspective, collections), personalization and recommendation (for selective and automatic dissemination of newly available documents by relying on dynamically produced user profiles), and collaborative work support (by implementing shared working spaces referencing users' own documents, collections, recommendations, related links, textual annotations, ratings, etc).(Candela, L., Straccia, U., 2003)(Avancini, H., Candela, L., Straccia, U., 2007).

A distinguishing typology of Digital Library Management System is represented by those dedicated to build a digital library by assembling a set of components. A notable example is represented by the *DelosDLMS*.(Ioannidis, Y., Milano, D., Scheck, H.J., Schuldt, H., 2008) This system has been developed in the framework of DELOS (Thanos, C., 2009) to integrate the various digital library services developed by DELOS members into a single working system. At the core of this system there resides an orchestrator that glues together the single entities as to implement the expected functionality in terms of chains of services calls. Another notable example of DLMS delivering a digital library by properly assembling existing components is represented by the series of tools (named *5SL*, *5SGraph*, and *5SGen*) relying on the *5S framework* (cf. Sec. 6) and designed by Gonçalves for modelling and semi-automatically customising digital library services. (Goncalves, M.A., 2004) 5SL is a declarative domain-specific language for digital library specification. 5SGraph is a domain specific visual digital library modelling tool whose output is a specification of a digital library in terms of the 5SL language. 5SGen is a component dedicated to the semi-automatic production of digital library components fulfilling the model of societies and scenarios expressed in terms of the 5SL language.

The rationale moving toward Digital Library Management Systems emerged quite naturally once the demand for digital libraries of various types started becoming diffuse and variegate. By analysing the development approaches characterising the first digital library systems – essentially based on from-scratch and ad-hoc development strategies – it becomes evident that they were neither proper nor sustainable if the goal is to serve production-oriented scenarios.(Ioannidis, Y., 2005) However, the lack of a common understanding of the functionality expected by digital libraries and the relative management systems led to the heterogeneous implementations described in this Chapter. Nowadays there is a relatively low number of systems that can be reasonably considered as "true" DLMSs, *i.e.* software systems equipped with management functionality supporting the development and operation of fully-fledged customised digital libraries. Despite these facts, the principle underlying them, i.e. *resources sharing*[23], is universally recognised as a valid one for reducing development and operational costs of digital libraries. Moreover, it has been the foundational principle leading to the notion of e-Infrastructure, a new digital library evolution frontier.

5. DIGITAL LIBRARIES EVOLUTION: INFRASTRUCTURES, VIRTUAL RESEARCH ENVIRONMENTS AND ECOSYSTEMS

Today scientific activities require collaboration among parties that are widely dispersed and autonomous. Collaboration is often cross-discipline and demands access to a variety of data and to specialized tools that support the analysis and processing of these data. If, in principle, digital libraries appear as potentially core enabling technologies for supporting such a new collaboration, in practice their application turns out to be too expensive to sustain.[24] Such a collaboration must rely on a wide range of heterogeneous and continuously evolving application resources, *i.e.* data and services, whose integration is very problematic as they are usually tailored to the specific requirements of the organisation that developed each one. Furthermore, the core functionality implemented by these applications (*e.g.* analysis, transformation, and extraction of knowledge from a large body of distributed and heterogeneous data) is computationally intensive and can rarely be sustained by individual organisations. Thus, setting up an appropriate collaboration framework is an expensive, time-consuming, and complex task that only few organisations can undertake in isolation.

To enable interoperability and uniform access to the heterogeneous wealth of available resources, new organizational patterns have been conceived, based on the notion of *e-Infrastructure*[25]. These approaches radically revolutionize the digital library organizational and development ideas by introducing a new paradigm which has also strong implication on all the digital library actors involved. According to such a paradigm, e-Infrastructures are technological solutions deployed and maintained operational by trusted organizations which guarantee their sustainability and the quality of the service offered to their users.

e-Infrastructures facilitate the realization of digital libraries to different extents. The majority of e-Infrastructures that have been created until now provide capabilities for the curation and access of domain specific resources. Typically they rely on a resource organizational model in which resource providers, which locally maintain and curate their own resources, agree on sharing them through under certain policies. The shared resources may range from publications, multimedia material, sensor and experimental data, to tools that manipulate these data, and computing and storage resources. A typical exemplar of this class of e-Infrastructures is the one built by the *IMPACT*[26] project. It offers mechanisms for aggregating, homogenizing, curating and accessing data stored in different archives of genomes and proteomes and maintains them for multiple consumption scenarios. Therefore, life science digital libraries can outsource the realisation of their information space of genomes and proteomes knowledge to the IMPACT e-Infrastructure, instead of implementing it and maintaining it operational. By exploiting this possibility the overall cost of the DLs is thus largely reduced. Another notable exemplar is the *GENESI-DR*[27] infrastructure, built by the homonymous project. It not only supports harmonization and uniform access to Earth Observation (EO) data, but also offers and mediates access to shared tools and computational facilities for generating EO products, like specialised information maps, resulting from the processing of the shared data.

A distinguished e-Infrastructure is the *D4Science e-Infrastructure*[28]. It adds a new facility to those offered by the e-Infrastructures described above. This new facility makes it possible to support the dynamic construction and maintenance of digital libraries, which in the context of this project are called *Virtual Research Environments* (*VREs*)(Assante, M., Candela, L., et al., 2008). VREs tailored to specific needs of a scientific scenario can be created and maintained for the time they are required, and dismissed when the community does not need them anymore (*e.g.* when a user community project comes to its end). The D4Science e-Infrastructure operates as a "broker" in a market of resources[29] accommodating the needs of resource providers and consumers. In the current version D4Science supports resource providers in "selling" their resources, and resource consumers, i.e. the scientific communities, in "buying" and orchestrating such resources to build their VRE applications.[30] The e-Infrastructure provides communities with logistic and technical aids for VREs building, maintenance, and monitoring in order to reduce as much as possible the human intervention and facilitate these tasks. Interactive tools are made available to support the selection of the resources to be included in these environments from the pool of the available assets. Once selected, these resources are organised and manipulated by the e-infrastructure in order to make the VRE operational, *e.g.* the services are deployed on specific servers, monitoring of these services is activated, reallocation is executed when needed. All these tasks are performed transparently to the users.

Each of the above e-Infrastructures offers a service over a set of resources. Despite this solution notably facilitates the construction of digital libraries that use the resources registered in the infrastructure, there is a growing evidence that the requirements raised by cross-disciplinary research may not be satisfied within the boundaries of a single e-infrastructure, regardless of how wide in geographical scale and large in aggregation capacity it may be. Rather the expectation is that collaboration will need to span across resources managed by multiple institutions, disciplines and countries, thus potentially ranging across multiple e-Infrastructures.

The response to this central requirement cannot certainly be the realization of a single global research infrastructure merging all the community- or discipline-oriented resources. Too many are the financial, organizational, and technological reasons that will ever prevent the realization of this solution. Recently, the research community as a whole has thus recognized that complying with today scientific enquiry still requires an additional step with respect to the e-Infrastructure solution experimented today. A more powerful and flexible organizational model capable of supporting interoperability and collaboration without forcing everyone to comply with a single model must be introduced. The new concept of *Knowledge Ecosystem* model has thus been proposed as a possible answer to this need. In a Knowledge Ecosystem single e-Infrastructures, although independent, are not isolated but dynamically interoperate and influence each other. They may share not only information, but also services needed to analyze and process the available information. In such ecosystem single digital libraries can offer a specific functionality to their user communities by relying on the support of other components of the ecosystem, thus the implementation of the functionality is outsourced to the ecosystem as a whole. The exploitation of the aggregated resources can then result in innovative applications made available to the communities served by the digital libraries which for reasons related to risk, cost, and scope are often excluded from the digital library roadmap for evolution. The realization of a Knowledge Ecosystem, which has just started to be investigated within the D4Science-II project[31], will require a considerable technological and organizational effort especially to deal with the interoperability issues, a very challenging issue also in this context.

Actually, the whole digital library development history and evolution, as presented in this Chapter, have evidenced how the multidisciplinary domain these innovative systems are requested to operate as well as the pragmatic and exploratory approaches adopted by the community for long time have concurred to characterise the digital library scope, its success stories as well as its drawbacks. Since the early times, digital library practitioners started borrowing solutions and approaches from other disciplines – including data management and library science – with the goal to experiment them to serve knowledge production needs. Synergies have been established between these disciplines and the digital library discipline. Substantial knowledge and experiences have been accumulated during this process. Unfortunately, despite the amount of evolution the field has reached and the enhancements it produced are tantamount to the evolution and enhancements of similar disciplines, a very limited effort has been dedicated to develop a foundational theory characterising the digital library domain. This is among the main reasons causing lack of success of some initiatives, hindering further digital library enhancements and convincing the practitioners on the need for renaming the field(Ioannidis, Y., 2005)(Atkins, D.E., Droegmeier, K.K., et al., 2003). However, the digital library is not completely lacking foundational oriented initiatives, as shown below.

6. A BIT OF FOUNDATIONS

Despite the life of Digital Libraries spans the last twenty years, a plethora of heterogeneous systems have been developed and classified under the digital library/digital repository umbrella. The development of a so large variety of systems, still ongoing, is not only due to the different application needs but also to the difficulty experimented in systematically describing, understanding, comparing and reusing digital libraries (and their constituents). This difficulty has its main root in the historical lack of foundations for them.

Among the first attempts to develop a digital library domain theory there is the *5S framework*. (Gonçalves, M.A., Fox, E.A., Watson, L.T., Kipp, N.A., 2004) It is based on five fundamental abstractions, *i.e. Streams*, *Structures*, *Spaces*, *Scenarios*, and *Societies*, to define digital libraries rigorously and usefully. Societies define how a Digital Library helps in satisfying the information needs of its users. Scenarios provide support for the definition and design of different kinds of services. Structures support the organisation of the information in usable and meaningful ways. Spaces deal with the presentation and access to information in usable and effective ways. Streams concern the communication and consumption of information by users. By having this model as foundational theory, a series of tools and systems have been designed and envisioned as to prove its effectiveness (Gonçalves, M.A., 2004).

Few years later, in the framework of the DELOS Network of Excellence, a very ambitious and challenging initiative started having the goal to provide the digital library community with a foundational, comprehensive and shared framework capable to capture the intrinsic nature of the various entities of the digital library universe. This initiative, by benefitting from the collective understanding developed by European research groups in the context of DELOS as well as from the international collaborations established in this framework, led to the *Digital Library Manifesto* (Candela, L., Castelli, D., et al., 2006) and to the *DELOS Digital Library Reference Model* (Candela, L., Castelli, D., et al., 2007). The former declaring the intentions, motives, overall plans and views of the initiative as well as introducing the main notions characterising the domain. The latter presenting the main concepts, axioms and relationships characterising the domain independently from specific standards, technologies, implementations, or other concrete details. Overall, the model distinguishes among three distinct notions of "systems" which are often confused in the literature: Digital Library; Digital Library System; and Digital Library Management System. These systems are characterized by a set of fundamental concepts belonging to six digital library specific domains, namely Content, User, Functionality, Quality, Policy, and Architecture. These systems support the operation of various actors playing four fundamental roles, namely End-User, DL Designer, DL System Administrator and DL Application Developer. Since December 2008, the development of the Reference Model is managed by the DL.org project (Castelli, D., Paker, S., 2009), an EU funded project promoting a consolidation and enhancement activity of this artefact on a scale involving the digital library community in the large.

Despite the lack of a foundational, well-established and universally accepted theory characterising the digital library domain, a lot of steps have been performed since the early stages and the initial conceptions of these systems supporting knowledge management. Also, the novelties introduced by the digital library field induced changes in our society, and its operational model.

7. IMPACT OF DIGITAL LIBRARIES

Probably the social and economical impacts digital libraries would have made on the library world were not recognized at the time when early digital libraries appeared. Certainly, the possibility of making scientific communication more effective and economic was in the mind of Paul Ginsparg while designing the arXiv system, even though not the perception that, after a few years from his primitive intuition, systems such as arXiv would have radically changed the way scientific communication had been conceived and put into practice. After a few years, however, Ginsparg passed from believing that "*in principle, the new electronic medium gives us the opportunity to reconsider many aspects of our current research communication, and researchers should take advantage of this opportunity to map out the ideal research communication medium of the future*"(Ginsparg, P., 1996) to the full awareness of the revolutionary changes arXiv was making in the communication of research information in many fields of physics. His awareness was well based. In fact, in 1997 the set of arXiv archives were serving over 50,000 users worldwide from over 100 countries, and processing many millions of electronic transactions per month. In some fields of physics, they had already supplanted traditional research journals as conveyers of both topical and archival research information. Thus Ginsparg could easily predict that "*the traditional model of funding publishing companies through research libraries (in turn funded by overheads on research grants) is unlikely to survive in the electronic realm*".(Ginsparg, P. 1997)

In giving an account of impacts of digital libraries on library world, this Section properly starts from the effects produced by arXiv archives in the physics community. But many other factors contributed to the changes digital library development gave start to (Borgman, C.L., 2007). Thus this "history" of impacts will continue proposing the economical crisis of libraries and the emergence of the Open Access Initiative as important economic and social factors strengthening the effects of digital technologies on issues such as business models, copyrights, etc., traditionally taken for granted in the practice of scholarly communication. The largest part of this Section, however, is dedicated to discuss how digital library evolution has made scientists to envision new way to work, and, in turns, how scientists' vision has moved digital libraries far beyond any connotation of the term "library". In this context, special attention is given to the new roles that both librarians and users are called to assume and to issues related to education for digital libraries.

One of the most important factors contributing to make changes desirable by library world certainly was the economical crisis of libraries themselves. In the latter 1990s, many financially pressed research libraries began to be poised for triage of their journal subscriptions. The majority of them began to consider the traditional model of journal subscription and book purchasing no more economically sustainable[32]. At the same time, the *Open Access Movement* emerged with the mission of disseminating knowledge widely and readily to society. In a conference convened by the Open Society Institute on December 2001, i.e., the Budapest Open Access Initiative, the goals of this movement were expressed in the opening sentence of the conference, as follows: "*An old tradition and a new technology have converged to make possible an unprecedented public good. The old tradition is academic scholars giving away the results of their research. ... The new technology is the Internet. Together, these have made it possible from everyone in the world to share knowledge freely and openly*". Four years later, the Berlin 3 Open Access Meeting[33] made new recommendations remedying the vagueness inherent in the Declaration's original wording about open access: "*In order to implement the Berlin Declaration institutions should: (1) Implement a policy to require their researchers to deposit a copy of all their published articles in an open access repository and (2) encourage their researchers to publish their research articles in open access journals where a suitable journal exists (and provide the support to enable that to happen)*".

If "electronic archives" opened the ways to substantial changes in scholarly communications, although originally thought for speeding dissemination only, its successors, i.e. institutional repositories, presented themselves as the tools for realizing open access goals, as can easily be understood. Moreover, the innovative functionality they have been provided with in the mean time were making them to emerge as a new strategy allowing "*universities to apply serious, systematic leverage to accelerate changes taking place in scholarship and scholarly communication*"(Lynch, C.A., 2003) and even "*rethinking*" it (Van de Sompel, H., 2004). This strategic role of repositories has recently been confirmed by the Association of Research Libraries[34].

In the early 2000s, the economic environment of libraries and the lively debate raised by Open Access movement[35] broke the delicate balance among the roles of authors, publisher and academic libraries, involving hot issues as intellectual property, copyright and the concept of "publishing" itself.

On the side of publishers, there were different reactions. The most important publisher of computer science literature, the Association for Computing Machinery (ACM), was the first to realize that a digital library of articles – and associated specialized services – had a greater chance of attracting scholars than simply providing subscriptions to printed and even electronic journals. Accordingly, it made an early strategic decision to orient its online development around a digital library rather than electronic journals. Discussing this choice, the ACM Deputy Director of Publications clearly put in evidence that socio-economic issues associated with the delivery of on-line content and services are as critical as sound technological implementation, if indeed not more so, and that ACM had distinguished itself by paying much attention to such issues. In particular, by giving own solution to the functioning of copyright law in a networked environment with the development of a new business model, mediating between free-and-easy dissemination and the demand for revenue.(Rous, B., 2001) No similar choice was taken by trade publishers, that, instead, inaugurated a business model dictating that libraries acquire access to bundle packages of journals, thus depriving libraries of their fundamental role of selector of quality materials.

Nowadays, the tensions between publishers and research libraries are far from being resolved(Shavell, S., 2009). However, some pacts of "no-belligerence" have been agreed, allowing authors to self-archive their research outputs into institutional repositories under certain conditions. This has given birth to different classes of publishers according to which copyrights conditions they are practicing.(SHERPA, n.d.) In the mean time, open access to outputs of publicly funded research is becoming a phenomenon more and more widespread[36,37]. Maybe this is the reason why someone affirms that a dialogue between publishers and librarians is possible. (Bowering, L., 2009)

In the same time digital libraries – or, more precisely, their primitive systems now called digital repositories – were restructuring the scholarly communication, the scientific community of the DELOS Thematic Network was working about a new vision for Digital Libraries, conceiving them as enabling "*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*". This vision was declared in the so-called "San Cassiano Report"[38] where also related socio economic issues were raised. In particular, beside the need of identifying business models for digital library operation and resolving copyright issues, the report recommended attention to how digital libraries could affect education and learning.

From that vision, digital libraries have made much progress in the direction of becoming universal knowledge repositories, making the wealth of material contained in libraries, museum, archives and any knowledge repository worldwide available. For this, Europeana (cf. Sec. 3) is a prominent exemplar. But the impact digital libraries are having on research libraries are of very particular nature, so that they are assuming connotations far beyond those inherent in the term "library".

Digital library evolution has made scientists to envision new ways how their work can be organized, and knowledge acquired, communicated and exploited. Scientists start dreaming integrated and collaborative working environments that by providing seamless access to the tools and the data they need offer an array of new research opportunities (Borgman, C.L., 2007). In their turn, advances in systems supporting e-research are inducing changes in the processes governing research activities in various fields as well as in what has to be conceived as end product of research itself. Datasets started becoming important research outcomes supplementing the traditional scholar communication objects and representing a valuable artefact for subsequent research. As described in Section 5, at the core of the current innovation there are *Virtual Research Environments*. Virtual Research Environments can be considered evolved versions of the current "research libraries", however they are revolutionizing traditional concepts with strong impacts on librarians and users also.

In such new environments, these traditional library stakeholders will possibly loose their specific connotations and assume many shared ones. Which role these "new librarians" will have can be hypothesized considering services that people working in VREs are engaged in. For example they are called to instruct the infrastructure on the specificities characterizing a resource including the policies governing its usage. This information serves to properly handle that resource. Different kinds of resources require that different information is specified[39]. The "new" librarians are also in charge to support resources ingestion[40], validate and approve them[41], as well as monitor their status[42]. Thus, for guaranteeing VRE operation a new specific profile could emerge, possibly a very new one that integrates users' specific competences (Candela, L., Castelli, D., Pagano, P., 2009).

VRE design and creation are other activities new librarians have to perform. These activities can be seen as an evolution of the more traditional ones played by librarians when supporting the library users in accessing the library content and services. Given the heterogeneity of the available resources and the complexity of the scientific processes that VREs may be called to support, these activities certainly require multiple expertise. In particular, the new librarians must fully understand the needs of the specific research communities asking for the VREs and the characteristics of the available resources. They must be capable of selecting the resources to be included, deciding their most appropriate configuration, functionality workflow, and so on.

This means that "new" librarians must have domain knowledge in the specific user community discipline, knowledge in information management, qualification in knowledge organization as well as be trained in IT.

The need for such an expertise should be supported by changes in digital library education. The need for information specialists capable to assist users in navigating complex information sources across heterogeneous repositories had already been evidenced in the end of nineties (Spink, A., Cool, C., 1999)(Schatz, B., Chen, H., 1999), however an effort to design a curriculum for DL education supporting teaching and learning about DL development and management was started only later, with the *Digital Libraries Curriculum Development project* (Pomerantz, J., Wildemuth, B.W., 2006). The curriculum modules, still under evaluation at the time of this writing, seem to fit well the students for dealing with new library's information objects – from multi-type, to multi-versioned multimedia documents – and organizational issues such as those inherent in the data-service provider paradigm. Organizational issues, in fact, will become more and more important as the infrastructure vision for federating repositories will advance, as the DRIVER confederation is demonstrating (Schmidt, B., Peters, D., 2008). But besides a managerial role, the DL evolving concepts and systems are calling for librarians and users as designers and operators (Candela, L., Castelli, D., et al., 2007) in *Virtual Research Environments,* as we have seen above. It is certainly not yet clear whether the entire VRE design, creation and maintenance process can be covered by a single professional. Certainly, however, innovative "librarians" profiles will have to emerge with complementary expertise from many disciplines (Lawton, F., 2009).

8. CONCLUSION

Digital libraries are undergoing a continuously evolving process, influencing all sectors where knowledge has to be created, stored, transmitted and used. This chapter has traced the history of digital library evolution through its fundamental steps, driven by how digital libraries could newly be conceived in connection with the availability of new technologies and the changing needs of the community of library users.

We have identified the first step in how early repository architecture was improved by the distributed ones, identifying their technical foundation in the Dienst system. The next important innovation was thinking about and dealing with interoperable technologies and frameworks, as succeeded in the Open Archive Initiative. Interoperability is the principle for content sharing and the basis on which all the worldwide digital libraries presently existing rely. The conception of Digital Library Management Systems represented the starting point for the spreading of digital libraries also in institutional environments not capable of supporting the cost of realizing a digital library from scratch. Since about 2000, Digital Library Management System projects such as Cyclades and Scholnet allowed for conceiving functionalities much different from the traditional ones, so that the envisioning of digital libraries as collaborative environments could emerge. Finally, e-Infrastructures, Virtual Research Environments and Ecosystems have been presented as the challenges the digital library research is facing today.

Social impacts is the last argument of the digital library's history as conceived in this chapter. It has primarily been concerned with impact on scholarship, with special attention on what has happened and is happening in the context of scientific research and development, for two main reasons. First, because digital libraries are regarded by national and international institutions as the central technology for the access, dissemination and preservation of scientific information (Council of the European Union, 2007). Second, because the actors in this context have particularly been forced to deal with the changing ways of making their profession - also contributing to tailoring its development, really. Accordingly, the impact on the role of librarians has received much attention, as librarians are destined to cover, in various degrees, a very great spectrum of the profiles needed by future research environments, as predicted in the Digital Library Reference Model(Candela, L., 2008).

But current events impose deep changes of the concept of research library itself, feeding the debate on which future is to be expected for digital libraries.

The library is transforming itself from a resource-based information system to a knowledge based service embedded into the research processes and collaborating with the researchers within the "knowledge ecosystem" that is being prospected as the needed future organizational pattern(Castelli, D., 2009). "*Knowledge organization, discovery, and experimentation are becoming a central part of research itself, not just passively supporting research, but actively or proactively stimulating, articulating, framing, guiding, and assessing research along the way right as the research is evolving. Research productivity in the future relies on this knowledge service infrastructure, and a new service mechanism is urgently needed to develop the infrastructure and to provide customized organizing, discovering, and computation services*" (Zhang, X., 2009). These are the messages, among others, recently sent in the second GRL2020 Asia in Taipei, Taiwan[43], where experts from around the world showcased best practices, case studies and pioneering work, with the aim of fostering innovative approaches supported by global research libraries.

ACKNOWLEDGMENT

This work is partially supported by the D4Science project, within FP7 of the European Commission, Theme INFRA-2007-1.2.2, Contract 212488 and by the DL.org Coordination and Support Action, within FP7 of the European Commission, Theme ICT-3-4-3, Contract 231551. Special thanks go to Maria Bruna Baldacci for her valuable help and suggestions in finalizing this chapter.

REFERENCES

Allen, P., Vaccaro, R., & Presutti, G. (2000) "ARTISTE: An Integrated Art Analysis and Navigation Environment". *Cultivate Interactive* (1), http://www.cultivate-int.org/issue1/artiste/

Altelman, K., Linema, E., & Pace, A. K. (2006). Toward a Twenty-First-Century Library Catalog. *Information Technology & Libraries*, *25*(3), 128–139.

Andreoni, A., Baldacci, M. B., Biagioni, S., Carlesi, C., Castelli, D., Pagano, P., & Peters, C. (1999). "Developing a European Technical Reference Digital Library. *Research and Advanced Technology for Digital Libraries, Proceedings of the Second European Conference on Digital Libraries, ECDL `99, Paris, France*, Lecture Notes in Computer Science, Springer, pp. 343-362

Arms, W. Y. (2001). *Digital Libraries*. The MIT Press.

Assante, M., Candela, L., Castelli, D., Frosini, L., Lelii, L., Manghi, P., et al. (2008). "An Extensible Virtual Digital Libraries Generator". *Research and Advanced Technology for Digital Libraries, Proceedings of the 12th European Conference on Digital Libraries, ECDL 2008, Aarhus, Denmark*, Lecture Notes in Computer Science, pp. 122-134

Atkins, D. E., Birmingham, W. P., Durfee, E. H., Glover, E. J., Mullen, T., & Rundensteiner, E. A. (1996). Toward Inquiry-Based Education Through Interacting Software Agents. *IEEE Computer*, *29*(5), 69–76.

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., & Messerschmitt, D. G. (2003). *Revolutionizing Science and Engineering through Cyberinfrastructure*. Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure.

Avancini, H., Candela, L., & Straccia, U. (2007). Recommenders in a personalized, collaborative digital library environment. *Journal of Intelligent Information Systems*, *28*(3), 253–283. doi:10.1007/s10844-006-0010-3

Belkin, N. (1999) "Undestanding and Supporting Multiple Information Seeking Behaviours in a Single Interface Framework". *Proceedings of the Eight Delos Workshop: User Interfaces in Digital Libraries*, ERCIM, pp. 11-18

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities http://oa.mpg.de/openaccess-berlin/berlindeclaration.html

Biagioni, S., Borbinha, J. L., Ferber, R., Hansen, P., Kapidakis, S., Kovacs, L., et al. (1998). "The ERCIM Technical Reference Digital Library". Research and Advanced Technology for Digital Libraries, Proceedings of the Second European Conference *on Digital Libraries, ECDL `98, Heraklion, Crete, Greece*, Lecture Notes in Computer Science, Springer, pp. 21-23

Borgman, C. L. (2007). *"The discontinuity of Scholarly Publishing". Scholarship in the Digital Age* (pp. 75–114). MIT Press.

Borgman, C. L. (2007). *"Data: Input and Output of Scholarship". Scholarship in the Digital Age*. MIT Press.

Bowering, L. (2009). "Publishers and Librarians: New Dialogues in Challenging Times". *Issues in Science and Technology Librarianship*, No. 56, http://www.istl.org/09-winter/viewpoint.html

Bush, V. (1945). As We May Think. *Atlantic Monthly*, *176*, 101–108.

CandelaL. (2008) op. Cit

Candela, L.; Castelli, D.; Ferro, N.; Ioannidis, Y.; Koutrika, G.; Meghini, C.; Pagano, P.; Ross, S.; Soergel, D.; Agosti, M.; Dobreva, M.; Katifori, V. & Schuldt, H. (2007). The DELOS Digital Library Reference Model - Foundations for Digital Libraries. DELOS: a Network of Excellence on Digital Libraries

Candela, L.; Castelli, D.; Ioannidis, Y.; Koutrika, Y.; Meghini, C.; Pagano, P.; Ross, S.; Schek, H. and Schuldt, H. (2006). "The Digital Library Manifesto". DELOS: a Network of Excellence on Digital Libraries

Candela, L., Castelli, D., & Pagano, P. (2003). "A Service for Supporting Virtual Views of Large Heterogeneous Digital Libraries". *Research and Advanced Technology for Digital Libraries, Proceedings of the 7th European Conference on Digital Libraries, ECDL 2003, Trondheim, Norway*, Lecture Notes in Computer Science, Springer, pp. 362-373

Candela, L., Castelli, D., & Pagano, P. (2009). On-demand Virtual Research Environments and the Changing Roles of Librarians. *Library Hi Tech*, *27*(2), 239–251. doi:10.1108/07378830910968191

Candela, L., Castelli, D., Pagano, P., & Simi, M. (2003). "From Heterogeneous Information Spaces to Virtual Documents". *Digital Libraries: Implementing Strategies and Sharing Experiences, 8th International Conference on Asian Digital Libraries, ICADL 2005, Bangkok, Thailand, December 12-15, 2005, Proceedings*, Lecture Notes in Computer Science, Springer, pp. 11-22

Candela, L., & Straccia, U. (2003). "The Personalized, Collaborative Digital Library Environment Cyclades and Its Collections Management". *Distributed Multimedia Information Retrieval, SIGIR 2003 Workshop on Distributed Information Retrieval, Toronto, Canada, August 1, 2003, Revised Selected and Invited Papers*, Lecture Notes in Computer Science, pp. 156-172

Castelli, D. (2009). "Creating a research library that preserves the past, present and curates the future". *GRL2020 Asia, Position Papers, 24-25 February 2009, Taipei, Taiwan*, pp. 11-12

Castelli, D., & Pagano, P. (2002). "OpenDLib: A Digital Library Service System". *Research and Advanced Technology for Digital Libraries, Proceedings of the 6th European Conference on Digital Libraries, ECDL 2002, Rome, Italy*, Lecture Notes in Computer Science, Springer, pp. 292-308

Castelli, D., & Pagano, P. (2003). "A System for Building Expandable Digital Libraries". *ACM/IEEE 2003 Joint Conference on Digital Libraries (JCDL 2003), 27-31 May 2003, Houston, Texas, USA, Proceedings*. IEEE Computer Society, pp. 335-345

Castelli, D., Pagano, P., & Straccia, U. (2001). "Scholnet and Cyclades: Extending the Role of Digital Libraries". *D-Lib Magazine*, 7(4), http://www.dlib.org/dlib/april01/04inbrief.html - CASTELLI

Castelli, D., & Parker, S. (2009). "DL.org: A Coordination Action on Digital Library Interoperability, Best Practices and Modelling Foundations". *ERCIM News*, 77, p. 65 CogPrints: Cognitive Sciences ePrint Archive. http://www.ukoln.ac.uk/services/elib/projects/cogprints/

CastelliD.ParkerS. (2009). op.cit

Christel, M., Kanade, T., Mauldin, M., Reddy, R., Sirbu, M., & Stevens, S.; Wactlar. H. (1995). Informedia Digital Video Library. *Communications of the ACM*, *38*(4), 57–58. doi:10.1145/205323.205337

Council of The European Union. (2007). "Scientific information in the digital age - Council conclusions". *Competitiveness (Internal Market, Industry and Research), 2832nd Council Meeting Press Release, Brussels, 22-23 November*, pp. 31-36

Council of the European Union. (2007). “Council Conclusions on scientific information in the digital age: access, dissemination and preservation”. *2832nd COMPETITIVENESS (Internal market, Industry and Research) Council meeting Brussels, 22 and 23 November 2007*

Crum, L. (1995). University of Michigan Digital Library Project. *Communications of the ACM*, *38*(4), 63–64. doi:10.1145/205323.205342

Davis, J. R., & Lagoze, C. (1995). Dienst: an architecture for distributed document libraries [DELOS Network of Excellence on Digital Libraries http://www.delos.info]. *Communications of the ACM*, *38*(4), 47. doi:10.1145/205323.205331

Davis, J. R., & Lagoze, C. (2000). NCSTRL: design and development of a globally distributed digital library. *Journal of the American Society for Information Science American Society for Information Science*, *51*(3), 273–280. doi:10.1002/(SICI)1097-4571(2000)51:3<273::AID-ASI6>3.0.CO;2-6

EPrints – Digital Repository Software Website http://www.eprints.org/software/

Fox, E. A., Akscyn, R. M., Furuta, R. K., & Legget, J. J. (1995). Digital Libraries. *Communications of the ACM*, *38*(4), 23–28. doi:10.1145/205323.205325

Fox, E. A., Eaton, J. L., McMillan, G., Kipp, N. A., Weiss, L., Arce, E., & Guyer, S. (1996). National Digital Library of Theses and Dissertations: A Scalable and Sustainable Approach to Unlock University Resources. *D-Lib Magazine*, (September): 1996. http://www.dlib.org/dlib/september96/theses/09fox.html.

Ginsparg, P. (1994). First Steps Towards Electronic Research Communication. *Computers in Physics*, *8*(4), 390–396.

Ginsparg, P. (1996). “Winners and Losers in the Global Research Village”. *Joint ICSU Press/UNESCOExpert Conference on ELECTRONIC PUBLISHING IN SCIENCE*. UNESCO, Paris, 19-23 February 1996. http://www.library.illinois.edu/icsu/ginsparg.htm

Ginsparg, P. (1997). Electronic research archives for physics. In Butterworth, I. (Ed.), *The Impact of Electronic Publishing on the Academic Community: An International Workshop Organized by the Academia Europaea and the Wenner-Gren Foundation* (pp. 32–43). London: Portland Press.

Gonçalves, M. A. (2004). “Streams, Structures, Spaces, Scenarios, and Societies (5S): A Formal Digital Library Framework and Its Applications”. PhD thesis, Virginia Polytechnic Institute and State University, November 2004

Gonçalves, M. A., Fox, E. A., Watson, L. T., & Kipp, N. A. (2004). Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries [TOIS]. *ACM Transactions on Information Systems*, *22*(2), 270–312.

Griffin, S., Peters, C., & Thanos, C. (2005). Toward the new-generation digital libraries: recommendations of the NSF/EU-DELOS working groups. *International Journal on Digital Libraries*, *5*(4), 253–254. doi:10.1007/s00799-004-0093-9

Ioannidis, Y. (2005). Digital libraries at a crossroad. *International Journal on Digital Libraries*, *5*(4), 255–265. doi:10.1007/s00799-004-0098-4

Ioannidis, Y., Maier, D., Abiteboul, S., Buneman, P., Davidson, S., & Fox, E. (2005). Digital library information-technology infrastructures. *International Journal on Digital Libraries*, *5*(4), 266–274. doi:10.1007/s00799-004-0094-8

Ioannidis, Y., Milano, D., Schek, H. J., & Schuldt, H. (2008). DelosDLMS. *International Journal on Digital Libraries*, *9*(2), 101–114. doi:10.1007/s00799-008-0044-y

Jones, S., & Manghi, P. (2009). "DRIVER: the Digital Repository Infrastructure Vision for European Research". *Zero-in e-Infrastructure News Magazine, EU FP7 Funded Project BELIEF-II*, 2, pp. 23-24

Kuil van der. A. and Feijen, M. (2004) "The Dawning of the Dutch Network of Digital Academic REpositories (DARE): A Shared Experience". *Ariadne* 41 http://www.ariadne.ac.uk/issue41/vanderkuil/

Lagoze, C., Fielding, D., & Payette, S. (1998) "Making Global Digital Libraries Work: Collection Services, Connectivity Regions, and Collection Views". *Proceedings of the 3rd ACM International Conference on Digital Libraries, June 23-26, 1998, Pittsburgh, PA, USA*, pp 134-143

Lagoze, C., Payette, S., Shin, E., & Wilper, C. (2006). Fedora: An Architecture for Complex Objects and their Relationships. *International Journal on Digital Libraries*, *6*(2), 124–138. doi:10.1007/s00799-005-0130-3

Lagoze, C., & Van de Sompel, H. (2001). "The open archives initiative: building a low-barrier interoperability framework". Proceedings of the first ACM/IEEE-CS Joint Conference on Digital Libraries, ACM Press, pp. 54-62

Lawton, F. (2009). "GRL2020 Position Paper". *GRL2020 Asia, Position Papers, 24-25 February 2009, Taipei, Taiwan*, pp. 43-45

Licklider, J. C. R. (1965). *Libraries of the Future*. Cambridge: The MIT Press.

Lynch, C. A. (2003). "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age". *ARL: A Bimonthly Report*, no. 226, pp. 1-7. http://www.arl.org/resources/pubs/br/br226/br226ir.shtml

Miller, P. (1999). Z39.50 for All. *Ariadne*, *21*, http://www.ariadne.ac.uk/issue21/z3950/.

NDLTD. Networked Digital Library of Theses and Dissertations. http://www.ndltd.org/

Paepcke, a.; Cousins, S. B.; Garcia-Molina, H.; Hassan, S.W.; Ketchpel, S.P.; Röscheisen, M.; Winograd, T. (1996). "Using Distributed Objects for Digital Library Interoperability". *IEEE Computer* 29(5), pp. 61-68

Payette, S., & Lagoze, C. (1998). "Flexible and Extensible Digital Object and Repository Architecture (FEDORA)". *Research and Advanced Technology for Digital Libraries, Proceedings of the Second European Conference on Digital Libraries, ECDL `98, Crete, Greece*, Lecture Notes in Computer Science, Springer, pp. 41-59

Payette, S., & Thornton, S. (2002). "The Mellon Fedora Project: Digital Library Architecture Meets XML and Web Services". *Research and Advanced Technology for Digital Libraries, Proceedings of the 6th European Conference on Digital Libraries, ECDL 2002, Rome, Italy*, Lecture Notes in Computer Science, Springer, pp. 406-421

Pomerantz, J., Wildemuth, B. W., Oh, S., Yang, S., & Fox, E. A. (2006). Digital Libraries Curriculum Development. *D-Lib Magazine*, *12*(7/8). http://www.dlib.org/dlib/july06/07inbrief.html#POMERANTZ.

RePEc: Research Papers in Economics. http://repec.org/

Rous, B. (2001). The ACM Digital Library. *Communications of the ACM*, *5*(44), 90–91. doi:10.1145/374308.374363

Savino, P., & Peters, C. (2004). ECHO: a digital library for historical film archives. *International Journal on Digital Libraries*, *4*(1), 3–7. doi:10.1007/s00799-003-0062-8

Schatz, B. (1995). Building the interspace: the Illinois Digital Library Project. *Communications of the ACM*, *38*(4), 62–63. doi:10.1145/205323.205341

Schatz, B., & Chen, H. (1996). Guest Editors' Introduction: Building Large-Scale Digital Libraries. *IEEE Computer*, *29*(5), 22–26.

Schatz, B., & Chen, H. (1999). Guest Editors' Introduction: Digital Libraries-Technological Advances and Social Impacts. *IEEE Computer*, *32*(2), 45–50.

Schatz, B., Mischo, W. H., Cole, T. W., Hardin, J. B., Bishop, A. P., & Chen, H. (1996). Federating Diverse Collections of Scientific Literature. *IEEE Computer*, *29*(5), 28–36.

Schmidt, B. and Peters, D. (2008). "DRIVER Repository Network Plan". *DRIVER Project Deliverable D2.1*

Shavell, S. (2009). "Should Copyright Of Academic Works Be Abolished?". Berkmman Center for Interne & Society at Harvard University. http://cyber.law.harvard.edu/node/5505

SHERPA – RoMEO. Publisher copyright policies & self-archiving. http://www.sherpa.ac.uk/romeo/

Smith, M., Barton, M., Bass, M., Branschofsky, M., McClellan, G., & Stuve, D. (2003). DSpace – An Open Source Dynamic Digital Repository. *D-Lib Magazine*, *9*(1). http://www.dlib.org/dlib/january03/smith/01smith.html. doi:10.1045/january2003-smith

Smith, R. S. (1996). A Digital Library for Geographically Referenced Materials. *IEEE Computer*, *29*(5), 54–60.

Smith, R. S., & Frew, T. (1995). Alexandria Digital Library. *Communications of the ACM*, *38*(4), 61–62. doi:10.1145/205323.205340

Spink, A., & Cool, C. (1999). Education for Digital Libraries. *D-Lib Magazine*, *5*(5). http://www.dlib.org/dlib/may99/05spink.html. doi:10.1045/may99-spink

Stanford Digital Libraries Group. (1995). The Stanford Digital Library Project. *Communications of the ACM*, *38*(4), 59–60. doi:10.1145/205323.205338

Straccia, U., & Thanos, C. (2004). An open collaborative virtual archive environment. *International Journal on Digital Libraries*, *4*(1), 23–24. doi:10.1007/s00799-003-0063-7

Tansley, R., Bass, M., & Smith, M. (2003). "DSpace as an Open Archival Information System: Current Status and Future Directions". *Research and Advanced Technology for Digital Libraries, Proceedings of the 7th European Conference on Digital Libraries, ECDL 2003, Trondheim, Norway*, Lecture Notes in Computer Science, Springer, pp. 446-460

Thanos, C. (2009). Digital Libraries: The pioneering Role of ERCIM in establishing this Research Field in Europe. *ERCIM News*, *77*, 10–11.

The Communications of the ACM (CACM), 38(4), 1995, a special issue on digital libraries contains short descriptions of many practical projects.

Van de Sompel, H., Krichel, T., Nelson, M. L., Hochstenbach, P., Lyapunov, V. M., & Maly, K. (2000). The UPS Prototype: An Experimental End-User Service across E-Print Archives. *D-Lib Magazine*, *6*(2). http://www.dlib.org/dlib/february00/vandesompel-ups/02vandesompel-ups.html. doi:10.1045/february2000-vandesompel-oai

Van de Sompel, H., & Lagoze, C. (2000). The Santa Fe Convention of the Open Archives Initiative. *D-Lib Magazine*, *6*(2). http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html. doi:10.1045/february2000-vandesompel-oai

Van de Sompel, H., Lagoze, C., Bekaert, J., Liu, X., Payette, S., & Warner, S. (2006). An Interoperable Fabric for Scholarly Value Chains. *D-Lib Magazine*, *12*(10). http://dlib.org/dlib/october06/vandesompel/10vandesompel.html. doi:10.1045/october2006-vandesompel

Van de Sompel, H., Payette, S., Erickson, J., Lagoze, C., & Warner, S. (2004). Rethinking Scholarly Communication – Building the Systems that Scholars Deserve. *D-Lib Magazine*, *9*(10). http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html.

Wactlar, H. D., Kanade, T., Smith, M. A., & Stevens, S. M. (1996). Intelligent Access to Digital Video: Informedia Project. *IEEE Computer*, *29*(5), 54–60.

Wilensky, R. (1995). UC Berkley's Digital Library Project. *Communications of the ACM*, *38*(4), 60. doi:10.1145/205323.205339

Wilensky, R. (1996). Toward Work-Centered Digital Information Services. *IEEE Computer*, *29*(5), 37–44.

Witten, I. H., & Bainbridge, D. (2002). *How to Build a Digital Library*. Elsevier Science Inc.

Woldering, B. (2004). The European Library: Integrated access to the national libraries of Europe. *Ariadne*, *38*, http://www.ariadne.ac.uk/issue38/woldering/.

Zia, L. L. (2001). The NSF National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) Program. *D-Lib Magazine*, *7*(11). http://www.dlib.org/dlib/november01/zia/11zia.html. doi:10.1045/november2001-zia

ENDNOTES

1 Associated with one or more regions ("footprints") on the surface of the Earth.

2 For instance, the Alexandria project has been continued by National Geospatial Digital Archive funded by the Library of Congress, University of California Santa Barbara, and Stanford University. See: http://www.ngda.org/

3 Actually, the DELOS Working Group was the result of an European Research Consortium for Informatics and Mathematics (ERCIM) initiative that in set up the its Working Group on Digital Libraries having as main objective to stimulate interest and to encourage collaboration between the ERCIM research teams beginning to be active in this field.

4 The *European Conference on Digital Libraries*. This annual conference rapidly became the main forum for the European DL community to present and discuss their research ideas and results.

5 Cross Language Evaluation Forum. http://www.clef-campaign.org/

6 Initiative for the Evaluation of XML Retrieval. http://inex.is.informatik.uni-duisburg.de/

7 The Cultural Heritage Applications Unit of the Information Society Directorate-General of the European Commission started having "digital libraries" among the research topics of the "DigiCult" (Digital Heritage and Cultural Content) area of 5th Framework Programme (FP5) for Research and Technological Development (1998–2002). In the course of the FP5 more than 100 projects in the DigiCult area were funded. This thematic priority area was also in the 6th Framework Programme (2002–2006) and is present in the current 7th Framework Programme (2007–2013).

8 The processes implementing the functionality and the content managed were residing on the same server.

9 This is very similar to the Vannevar Bush "dream" expressed with the knowledge and the potentialities of today's society.

10 Z39.50 Maintenance Agency web page http://www.loc.gov/z3950/agency/

11 Dublin Core is the common name for the Dublin Core Metadata Element Set, a vocabulary of fifteen properties for use in resource description developed by the Dublin Core Metadata Initiative (http://dublincore.org).

Because of its simplicity and "core" nature, i.e. its elements are broad and generic, it has been largely used to describe a wide range of resources.

12 NARCIS (http://www.narcis.info/) provides access to more than 240,000 scientific publications (the majority of them are open access publications), more than 6,000 data sets, and information on researchers (expertise), research projects and research institutes in the Netherlands.

13 The NSDL program held its first formal funding cycle during 2000. From 2000 onward, over 200 projects have been funded to create collections, services, and tools for teacher and learners at all levels, and perform targeted research in digital libraries and their application to education. The NSDL program is an unusual program for NSF in that its projects are engaged in building an enterprise much larger than the object of any one grant. As of October 2008, the NSDL transitioned to a new phase of development and organization, with the granting of awards for the NSDL Resource Center (RC), and Technical Network Services (TNS). http://nsdl.org/about/?pager=organization

14 Europeana http://www.europeana.eu

15 Europe's Information Society – i2010: Digital Libraries Initiative http://ec.europa.eu/information_society/activities/digital_libraries/index_en.htm

16 Digital Repository Infrastructure Vision for European Research. www.driver-community.eu

17 DRIVER Confederation is the name chosen for this organization at the time of the writing of this Chapter. The name may be changed to reflect the more international level that the Confederation is starting to cover including the U.S., Canada, Latin America, China, Japan, India and Africa.

18 D-NET. http://www.driver-repository.eu/D-NET_release

19 D-Net is equipped with a set of tools that enable to register the repositories willing to share their content, check a number of quality parameters for these repositories, harvest through OAI-PMH, clean and integrate metadata records according to target metadata record formats. The resulting Information Space can then be accessed via an arbitrary number of DL applications built by service providers. D-Net also provide number of predefined and configurable services, such as Recommendation, Collection, Browsing, and User Interfaces that can be exploited by service providers in building their specific application.

20 Open Archives Initiatives – Object Reuse and Exchange http://www.openarchives.org/ore

21 A list of known DSpace instances is reported in the DSpace website (http://www.dspace.org/index.php/DSpace-Instances/Repository-List.html). In August 2009 this list contains more than 600 repositories.

22 DuraSpace. http://duraspace.org

23 The notion of "resource" has to be intended with the most abstract and generic meaning here as to potentially capture any entity in the Digital Library universe. Thus "sharing" should be realised on content resources as well as on functionality, user and any other resource having a value in a system different from the one it was been built for.

24 Existing technologies for content sharing (cf. Sec. 3) and Digital Library Management Systems (cf. Sec. 4) are valid approaches toward the realisation of such systems. Unfortunately, they are not yet mature enough to deal with the plethora of issues arising while dealing with the very variegated scenario eScience is posing.

25 The term 'e-Infrastructure' refers to research environment in which all researchers – whether working in the context of their home institutions or in national or multinational

scientific initiatives – have shared access to unique or distributed scientific facilities (including data, instruments, computing and communications), regardless of their type and location in the world. http://cordis.europa.eu/fp7/ict/e-infrastructure/

26 IMproving Protein Annotation through Coordination and Technology, http://www.ebi.ac.uk/impact/page.php

27 Ground European Network for Earth Science Interoperations - Digital Repositories, http://www.genesi-dr.eu/

28 DIstributed colLaboratories Infrastructure on Grid ENabled Technology for Science, http://www.d4science.eu

29 *Resources* here are intended as shareable generic entities, physical (*e.g.* storage and computing resources) or digital (*e.g.* software, processes, data), that can interact with other resources to synergistically provide some functions serving their clients, either humans or automatic systems.

30 Selling is supported through the publishing of resources according to the policies established by their owners. The proprietary formats and protocols used by these resources are transformed into common ones by the e-infrastructure services and facilities so that their seamless consumption is enabled. The pool of resources shared by third-party providers is enriched by a set of service resources, *i.e.* software units which deliver generic digital library functions, like retrieval, access, annotation of content and creation of new one. This pool of functionality which constitutes a core part in the majority of the VREs, can be used as any other public resource by exploiting available physical resources, implemented and made available by the e-Infrastructure itself.

31 D4Science-II is a project recently founded, whose starting date is October 2009.

32 The crisis is still ongoing, as documented in the "ARL Statement to Scholarly Publishers on the Global Economic Crisis" issued by the Association of Research Libraries in 2009 and announced in the Association's Press Release as "The Global Economic Crisis and Its Effect on Publishing and Library Subscriptions: ARL Issues Statement to Scholarly Publishers and Vendors" http://www.arl.org/news/pr/econ-crisis-19feb09.shtml

33 Berlin 3 Open Access: Progress in Implementing the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Feb 28th - Mar 1st, 2005, University of Southampton, UK. http://www.eprints.org/events/berlin3/outcomes.html

34 Association of Research Libraries. (2009). "The Research Library' Role in Digital Repository Services. Final report of the ARL Digital Repository Issues Task Force". Association of Research Libraries, http://www.arl.org/bm~doc/repository-services-report.pdf

35 A comprehensive overview of the debated issues regarding Open Access can be found in the dedicated web site (http://www.earlham.edu/~peters/fos/overview.htm) maintained by Peter Suber, one of the promoter of the movement.

36 In May 2009 SHERPA announced that its service RoMEO was listing 600 publisher policies on self-archiving.

37 The ROARMAP (Registry of Open Access Repository Material Archiving Policies), accessed in August 2009, listed 1436 open access repositories distributed worldwide. http://roar.eprints.org/

38 "Digital Libraries: future directions for a European research programme". DELOS Brainstorming Report, San Cassiano, Italy, June 2001. http://delos-noe.isti.cnr.it/activities/researchforum/Brainstorming/1st-ws.html

39 For instance, if the resource is a web service implementing a specific functionality

its URL has to be provided, if the resource is a data source both a characterization of its content and the protocol governing the access to it must be given.

[40] The e-Infrastructure needs to enrich the pool of resources explicitly specified at registration time. Librarians will guide the process complementing this pool with additional resources facilitating the exploitation of the initial ones. For instance, in the case of data sources, metadata collections in specific schemas can be generated, new collections of information objects resulting from original data aggregation and manipulation can be produced, different indices supporting data discovery can be automatically generated.

[41] Librarians are requested to analyze the characteristics of the registered resources and decide whether these resources are entitled to partake the infrastructure or not.

[42] The data needed to monitor the resource status are per resource, i.e. the status of different resources is characterized by different aspects. For instance, the status of a web service includes its workload, the status of a data source includes the number of information objects it contains.

[43] GRL2020 Asia, 2009. http://www.grl2020.net/index.php/review