
Event Report
June 13, 2007

Joint DELOS – NSDL Summer School

Digital Libraries for the Digital Librarian: Making the Journey from Traditional to Digital Libraries

28 May - 2 June 2007
Settignano, Florence (Italy)

**Prepared for the
Humanities Advanced Technology and Information Institute (HATII),
University of Glasgow**

By

Leonidas Konstantelos
HATII, University of Glasgow UK
L.Konstantelos@hatii.arts.gla.ac.uk

This document reports on the Joint DELOS – NSDL Summer School, which took place on Settignano, Italy, from 28th May to 2nd June 2007. The Summer School was entitled, “Digital Libraries, Making the Journey from Traditional to Digital Libraries” and aimed at informing attendees about the technologies and organisational issues involved in shifting from a traditional organisational structure to a Digital Library, “illustrating criteria and methods that exploit the strengths of digital libraries in a socio-economic and interdisciplinary manner.”¹ The structure of the lectures followed a top-down approach, starting with fundamental practices of traditional libraries then progressively moving toward the information dissemination cycle in the digital realm. In addition, work groups were intended to test the knowledge acquired in terms of the impact of the digital medium in information deployment, management and underlying technology. The school began on the evening of 27th June with a welcome reception at Villa Morghen. Lectures started the following morning and included sixteen presentations in total and two visits to cultural institutions in Florence.

The opening session on 28th May started with a speech by Prof. Paolo Galluzzi, Director of the Istituto e Museo Nazionale di Storia della Scienza (IMSS) in Florence. Galluzzi commenced with a critique of the term “digital library”, stating that it is an ambiguous name that does not encompass the many implications of this technology on the knowledge society. Instead, he suggested that the notion of “hybrid library” expressed by Rusbridge² is more suitable for describing the transition state of the library, which is currently neither fully analogue nor fully digital. Galluzzi explained that DLs extend the traditional repository of books and manuscripts, by incorporating a combination of – often incoherent – material types that together create a collective, unified learning experience. The evolution of the ALM domain (Archives, Libraries and Museums) has historically produced fragmentation in the Culture Heritage disciplines (for instance, collections of scientific material have been separated from artistic ones), which must not be carried in the migration from analogue to digital format.

With respect to the History of Science and Technology, hybrid libraries aid in making scientific and technological heritage widely accessible. The project entitled “Galileotheke@” is an integrated digital repository of

1 http://www.delos.info/index.php?option=com_content&task=view&id=566&Itemid=305

2 Rusbridge, C. (1998) Towards the Hybrid Library. *D-Lib Magazine*, July/August 1998.

Online: <http://www.dlib.org/dlib/july98/rusbridge/07rusbridge.html>

Galileian resources that incorporates a wide range of hybrid, diverse and interconnected material. Galileothek@ bundles text, images, 3D models of research tools, lexicons of terms and experiments, bibliographic records, documents and iconography with a virtual museum. Galluzzi exhibited through examples how horizontal access to different archives of the resource offer a continuous browsing experience. Building on this project, he emphasised the demand for semantically important information retrieval not incorporated in search engines, where cultural and domain-specific context is absent.

The last part of the talk focused on the presentation of Galileo's Blog, a resource built according to the Web 2.0 standards and based on open-source software to support the functions of the blog. According to Galluzzi, Galileo's Blog marks the passage from taxonomies to folksonomies³, stimulates user participation with semantically structured layers of comments, desiderata and views, converges the expertise of different professionals under the web infrastructure and promotes the ex-post evaluation of community material. Galluzzi concluded that the Web offers a fertile terrain for anthropological, sociological and cultural analysis; blogging challenges traditional channels of scholarly communication, calls for a revision in cultural production practices and presents a promising transition to the new media environment of information sharing.

The next lecture, led by Vittore Casarosa of the Information Science and Technologies Institute - Italian National Research Council (ISTI-CNR), offered a comprehensive introduction to Digital Libraries and related technologies. Casarosa first introduced the DELOS Network of Excellence, defining its activities as coordination of research, publications, dissemination and visibility, workshops, meetings, tutorials/summer schools and researcher exchanges. He further presented the objectives and visions of DELOS for the digital library, which revolve around universal, multi-modal access to cultural material and exchange of expertise.

Casarosa accepted that a technology as versatile and interdisciplinary as digital libraries cannot be contained in one single definition, but largely depends on the scholarly orientation from which a DL is viewed. In the most abstract form, digital libraries can be perceived as a combination of content and services/functionality. He argued that this technology is better

³ The term is attributed to Thomas Vanderwal and describes user generated taxonomies.

understood through a conceptual framework that establishes a common vocabulary, defines entities and acts as a yardstick for DL systems. The conceptual model presented consists of three interconnected components: a Digital Library Management System (DLMS) that takes input from domain experts, administrators and application developers; a DL system responsible for establishing connections among various nodes of a distributed architecture; and a Digital Library that is materialised by user interfaces, services and information output.

According to Casarosa, DELOS has focused on the DLMS component. The vision for future libraries is to move beyond the content-centric, isolated systems that are targeted for static storage, and deliver applications that are person-centric, designed for communication/collaboration within and among communities of users, through a global distributed interacting infrastructure. In this sense, a generic DLMS technology to accommodate different content and contexts can help overcome the current practice of environment-specific repeated efforts. The DELOS research directions are founded on these considerations and cover a wide range of activities, from applications and user requirements to system design, evaluation, curation and preservation of digital material. Casarosa centred on the importance of surmounting semantic interoperability, because a problem in representing knowledge in computer software is the lack of understanding how human beings perceive and process information (for instance, in visual representations).

In the final part of the lecture, Casarosa tried to abstract the concept of the “digital library” in an effort to disambiguate its meaning. An emerging term is 'knowledge commons' but most of the participants found it is even more difficult to grasp primarily as a concept. The body concluded that onomatology is subjective and should not be viewed as the cornerstone of digital library development; instead, the objective should be the level and quality of research that enables the actualisation of the DL vision – or as Casarosa put it, the process of “getting...there”.

The morning session closed with an inspiring speech from Kaye Howe, Executive Director of the National Science Digital Library (NSDL; University of Colorado, US). Howe highlighted the educational outcome of digital libraries, but cautioned about the power of education in imposing uniformity or variability of beliefs and ideas in societies. She presented their strategy at NSDL to reanimate K12 students' interest into science, by leveraging trusted systems of teaching, creating a domain-

specific, but generic and tailorable infrastructure, and by promoting sharing of knowledge / expertise. Their goal is to provide teachers with the resources needed to motivate students toward scientific domains; such material must de facto be curated and context-centric. Based on the Intellectual Context Maps by AAAS⁴, the NSDL allows teachers to research a topic to a level of granularity that is based on their particular needs.

In order to understand these requirements, a 4-year project on assessing user needs has been conducted, with educators, children and parents as the survey participants. Howe briefly reported on the results, which revealed two interesting facts: discrepancies among different social groups were less acute than speculated; and classroom participation is becoming irrelevant to learning, as the Internet becomes a growing source of information. In the light of these findings, Howe emphasised the impetus for DLs to represent a cognitive tool that provides the means to trustworthy, scientific material.

In the afternoon session, Laura Campbell of the Library of Congress (LOC; US) presented the institution's strategy for digital collections. The first part of her talk introduced the Library of Congress through historical facts and figures. Campbell then explained how the digital collections were actually built. At the initial stage, primary source material of maximum usefulness to educators was identified through collaboration with curators, domain- and educational experts. The collections have been redesigned to match the specific needs of teachers and students; to this end, the Learning Page resource accommodates material to support the 'Teaching with Primary Sources' programme.

The success of these initiatives led to the understanding that a trusted repository was needed to ensure long-term availability of the material, especially in cases of 'born-digital' information. With respect to the latter, the amount of material produced is daunting; social and technical issues influenced decision-making on a preservation methodology. Collaboration between LOC with other institutions of the federal government and the private sector was mandated by the US Congress, thus forming a network of partners - the National Digital Information Infrastructure and Preservation Program (NDIIP) - to tackle the problem of digital preservation. Campbell reported on the success of this

4 American Association for the Advancement of Science (<http://www.aaas.org/>)

collaboration, which allowed for more than 66 terabytes of at-risk material to be saved and preserved. She concluded her talk by demonstrating the future investments of the NDIIP in three areas: creation of content in the public sector (films, digital art, video games); preservation of important US state records; and support of current partners in continuing their work.

The final session of the day was introduced by Ragnar Nordlie of the Oslo University College (Norway). His talk looked at the user perspective between the traditional and the digital library. In physical libraries, the values of retrieval, quality, reliability, local relevance and seamless access are highly important. When moving to the digital realm, Nordlie posed the question of whether users will exhibit the same behaviour in accessing the online resources, in terms of purpose, material and metadata. In addition, he questioned whether the digital library can be managed by employing traditional library methodology.

Setting off from these considerations, he presented the experience of the Norwegian National Library digitisation effort; the NNL is aiming high at digitising its full collections in a time span of 15 years. A trial project to digitise 20000 items was launched in 2007, which highlighted the differences between the traditional library model and the user search behaviour. Summarising these findings, Nordlie suggested that users turn to the library as a last resort to access specific information, mainly because Google is perceived as an ideal system model and there is reluctance in exerting effort to understand advanced search facilities and adapt one's vocabulary to that of the library system. He concluded that recurring problems in such initiatives are (1) ensuring accessibility to the digitised material, since negotiations with rights management organisations can hinder free and timely access; and (2) keeping pace with current technological developments.

At the end of the first day, Pat Dixon of Northumbria University (UK) gave a brief introduction into Project Management and assigned the work group activity, which involved attendees working in groups to design the ideal digital library of the future by employing the Project Management principles as a guide.

The next morning, Donatella Castelli of ISTI-CNR and Dean Krafft of Cornell University (US) led a lecture on organising the digital library. Castelli presented the introductory concepts concerning digital libraries,

particularly in terms of the conceptual model presented the day before by Casarosa. The discussion started with definitions of the Digital Library, the DLMS and the DL actors; with respect to the latter, Castelli identified four types of actors: end-users, designers (which includes domain experts), system administrators and application developers. The focus then moved to the functions implemented by a DL System (DLS) and the application framework that provides the run-time environment for the DL enabling components to function and interact. This functionality was further analysed into mediation, information access, information management, personalisation, collaboration between repositories in a distributed environment and presentation.

In the second part of the lecture, Krafft explained the implications of the transition from a digital collection to a digital library; in particular, he referred to heterogeneity of systems in a distributed application infrastructure, the level of human intervention and the 'perfection' sought after. An identified problem lies in the combination of different metadata standards from many collections in a uniform manner. To this end, he presented the Open Archives Initiative (OAI) Protocol for Metadata Harvesting (OAI-PMH). OAI-PMH is a protocol for aggregating collection metadata of any XML schema for batch mode, thus allowing a record to be described in multiple ways. Although it requires Dublin Core for interoperability, the protocol has been extended to include compound objects (such as METS and DIDL).

Krafft moved on to showcase an application of the OAI-PMH in the context of the NSDL and identified the strengths and weaknesses of an automated metadata harvesting and aggregation process. In this spirit, the notion of the Semantic Digital Library was introduced as an informational space that integrates material descriptions based on different metadata standards, provides interoperability between systems and delivers interfaces empowered by semantics. In the last part of the lecture, Krafft demonstrated the role of DLs as collaborative social spaces, by utilising the NSDL wiki, tagging, bookmarking and recommendation functionalities as an example. The session concluded with a demo of DRIVER (Digital Repository Infrastructure Vision for European Research) presented by Paolo Manghi of ISTI-CNR. The demo convincingly exhibited how the DRIVER test-bed allows for a distributed, service-based, sustainable environment that is built upon web services, layer components and the DELOS DLS/DLMS.

In the afternoon session, a visit to the Kunsthistorisches Institut in Florenz – Max Planck Institut was planned. During the visit, the attendees had the opportunity to view a presentation of the institute's scope and research activities and gain access to the collections.

In the morning session of the third day, Paul Weston of the University of Pavia (Italy) led a lecture on 'Describing Information' in the context of a digital repository. The discussion started with a short history of the term "metadata", as its use passed from database systems and the World Wide Web to the library environment; in the case of the latter, the term applies to any format scheme of resource description for any type of object. Weston identified three bands in a typology of metadata formats where the simpler formats are used for full indexing and are automatically extracted by web search services, structured formats (e.g. Dublin Core, RFC 1807) offer an intelligible internal structure for non-specialists, and rich formats (MARC, EAD) provide a structure for more descriptive information, both for resource discovery and for the larger task of documenting objects or collections of objects. The lecture then focused on the organic role of metadata for cataloguing, information discovery and organisation of electronic resources. Weston explained the different ways in which metadata can be stored and pinpointed that the selection of a metadata schema depends on the type of information that an organisation wishes to provide, the mode of access and the degree of granularity required. Towards the end of his presentation, Weston very briefly (due to lack of time) presented two metadata projects: the Dublin Core (now DCMI) and the Linking and Exploring Authority Files (LEAF) project.

The next presentation was led by David Millman of Columbia University (US) and centred on identity management in distributed services. The discussion highlighted the importance of incorporating identity control in a distributed environment, such as that of a digital library, for restricting unauthorised access, providing personalised functions and utilising log monitoring for evaluation and usage metrics. Although enforcing a uniform policy in a system that actually comprises of a number of individual institutions is far from straightforward, the dimensions of trust must be expressed in a formal specification that identifies the trade-offs between the value/risk of the services, privacy and relationships with partners and users. Millman presented the Level of Assurance (LOA) standards process, a trust management tool to assess the strength of identity assertion. The talk ended with three useful recommendations for scaling identity, in terms of an identity proof mechanism, data profiling

and decisions on authorisation. A vivid discussion followed among the school participants about the interference of such policies in the free, seamless access to information promised by DLs.

In the afternoon lecture, Carlo Meghini (ISTI-CNR, Italy) talked about the issue of information access in Digital libraries. After clarifying the concepts to be discussed, Meghini distinguished two information access methods: discovery via content and discovery via associations. This led the discussion to the issues of “best match” and “exact match” approaches and how these information access techniques are moulded depending on the type of descriptions employed. Meghini provided a comprehensive account about the concepts of taxonomies, folksonomies, ontologies and semantic interoperability; although these are complex ideas to explicate within a single lecture, participants enthusiastically claimed that Meghini's approach was more than enlightening. The focus then moved to personalised information access and a comparison between quantitative and qualitative personalisation. In the last part of the session, Meghini outlined the challenges for distributed DLs in gathering information descriptions, managing query evaluation and efficiently dealing with syntactic and semantic interoperability. He concluded that the field of information access provides fruitful ground for further research, as the ever-expanding *knowledge* poses difficulties in collection and representation while the many faces of semantic interoperability assimilate a Babel Tower situation.

Thursday morning's session started with Sandy Payette of the Cornell University (USA), the founder and co-director of the internationally acclaimed Fedora project; Fedora (Flexible Extensible Digital Object Repository Architecture) is an open-source project that provides sophisticated software for the development of digital repositories⁵. The talk first looked at the complex issue of user needs' evolution and the implications of Web 2.0 for understanding the environment that future scholars will be working in. Payette cautioned in a number of occasions during her talk that the DL community must plan now for technologies to be service-oriented, flexible and decomposed to allow for smoother integration into the upcoming Web status. In order to support this view, she presented the Plos One and NSDL 2.0 Blog projects that utilise the Semantic Web technologies integrated into Fedora. She argued that the Resource Description Framework (RDF) and community-based ontologies

5 <http://www.fedora.info>

are naturally extensible and can therefore enable semantic interoperability and complex queries. She then exhibited how digital objects are instantiated in Fedora by use of RDF, permitting 'compositional semantics'. She concluded with examples of DL projects from various domains that have successfully employed the Fedora infrastructure and reiterated the impetus for technology to evolve alongside with user needs.

In the following lecture, John Akeroyd of University College London (UK) provided a detailed introduction to the role of digital libraries in e-learning. Akeroyd first described the advantages and perceived disadvantages of e-learning, making particular reference to the British educational system. He then set off to define the Virtual Learning Environment (VLE) and the rationale behind its integration with a digital library system. He suggested that a Managed Learning Environment (MLE) could be more suitable, as it offers links to both management systems of an academic institution (such as student records) and libraries, portfolios, regulations and handbooks. Akeroyd displayed progress made in this field through the Angel and Easel projects. He emphasised on the need for e-learning standards and pedagogic metadata, especially in the light of new developments that demand Universities to reflect the changing agenda of a distributed student community. He concluded with a discussion of how the Web 2.0 innovations set the path for Personal Learning Environments, in which the user possesses control over instrumentation and the student-University relationship radically changes.

The afternoon session involved a visit to the Institute and Museum of History of Science in Florence, where the foundations of the Galileothek@ project were supposed to be presented. Instead, Andrea Scotti of the Fondazione Rinascimento Digitale (Italy) gave a brief presentation of the Pinakes project and the foundation's work on systems that utilise ontologies for the Humanities. The visit to the museum was limited to a guided tour of a temporary exhibition about sundials, which -though irrelevant to the school's subject- was quite interesting.

The last day of the school commenced with a joint lecture by Anna Maria Tammaro of the University of Parma (Italy) and Mary Marlinio of the Digital Learning Science (US). The lecture looked at the issue of evaluation in digital libraries. Tammaro offered a brief yet detailed introduction to the subject, outlining the fundamental purpose of the process for accountability of the resources, effectiveness of services and

impact on information dissemination. Tamaro explained that evaluation should be viewed as a cycle of interrelated processes revolving around the objectives of the digital library. She distinguished between content-centred, system-centred and user-centred evaluation, presented methods to assess resources and concluded with the notion that future resource evaluations might be self-employed by the users of different repositories.

Marlino built on the previous discussion and introduced three important pitfalls when evaluating resources: the tendency to perceive material description as the primary goal; the reluctance to embrace complexity, but instead oversimplify the process; and the need to render proper judgement. To avoid such pitfalls, the nature of the evaluation must be specified, either it be formative, summative, effectiveness or impact evaluation. Marlino argued that evaluation can be a political activity, especially when the goal is to understand and encompass the impact of multiple publics into the project. An interesting point made is that (in Marlino's words) "evaluation doesn't *prove* anything"; even though evidence in terms of survey data is a source of deliberation, people are responsible for the impact of decision-making. As a case study to the concepts reviewed, Marlino presented the Digital Library for Earth System Education (DLESE), a project based on a user-centred design process for which ongoing evaluation (currently in the 7th year) has taken place.

The morning session closed with a presentation by Carl Lagoze of the Cornell University (US) on scholarly communication in the 21st century. The issue was introduced by first explaining what is meant by 'scholarly communication', the functions of this system and the drawbacks of current practices. Lagoze argued that scholarly communication is a "system under pressure", as electronic distribution of material has brought changes in economics (e.g. publishing/printing costs, costly subscriptions to journals) and the advent of Web 2.0 creates speculations of how the next-generation of scholars will seek information. Lagoze suggested that object re-use and exchange can revolutionise the function of the scholarly communication system; in this context, he presented the OAI-ORE effort to "develop, identify, and profile extensible standards and protocols to allow repositories, agents, and services to interoperate in the context of use and reuse of compound digital objects beyond the boundaries of the holding repositories." He concluded that the Web and repositories of all kinds form the building blocks for new knowledge networks, which necessitate commonly accepted models and protocols for

information exchange – especially in the case of complex, composite and interrelated information units.

During the final session of the summer school, the work groups presented their findings. Although time constraints did not allow the original assignment to be carried out as planned; instead, a new set of specifications instructed each group to present the debate between its members, in terms of decision-making. Some interesting concepts were introduced, especially as regards the definition of a digital library, the potential methods for creating devoted, 'passionate' users of a DL system and techniques to avoid compromise in information access.

On the whole, the DELOS Summer School was successful, as the majority of the participants claimed to have acquired a better understanding of the purpose and organisation of digital libraries. Among those attendees representing cultural institutions, there was a vivid interest toward focusing on the development of a DL system to fulfill institutional goals and leverage the dissemination of cultural material in their collections to a wider audience. This fact alone not only marks the accomplishment of the Summer School, but also justifies the systematic efforts of the digital library community to surpass the barriers of traditional practices in the ALM domain and accommodate Culture Heritage Institutions with contemporary, state-of-the-art solutions.