

Scalable, Interoperable Digital Library Infrastructure
Ronald L. Larsen, University of Pittsburgh

The prospects for a Maghreb Digital Library for Education, Science & Culture raises tantalizing opportunities for both those who conduct advanced research related to digital libraries and those whose interest is in documents and artifacts from or about this vital region. In this brief paper, I will draw on the results of the 2003 NSF workshop on research directions for digital libraries (<http://www.sis.pitt.edu/~dlwkschop/papers.html>) held in Chatham, Massachusetts to suggest topics of potential importance to a Maghreb Digital Library for Education, Science & Culture (MDLESC).

The Chatham report recommended digital library research focused on:

- Increasing the scope and scale of information resources and services;
- Employing context at the individual, community, and societal levels to improve performance;
- Developing algorithms and strategies for transforming data into actionable information;
- Demonstrating the integration of information spaces into everyday life; and
- Improving availability, accessibility, and, thereby, productivity.

While research in these areas could ultimately impact favorably on a MDLESC, the recommendations for development of *scalable, interoperable digital library infrastructure* seem more relevant to the immediate interests of a MDLESC. The infrastructure program recommended in the Chatham report includes:

- Acquisition of new information resources;
- Effective access mechanisms that span media type, mode, and language;
- Facilities to leverage the utilization of humankind's knowledge resources;
- Assured stewardship over humanity's scholarly and cultural legacy; and
- Efficient and accountable management of systems, services, and resources.

Indeed, reflecting on these recommendations, they seem particularly appropriate to a MDLESC.

Acquisition

Millennia of cultural artifacts and documents likely make the Maghreb one of the richest and most diverse sources of materials to challenge the capacity and capability of digital libraries. Beyond the materials themselves, current advances in laser 3-D imaging and the application of medical imaging (e.g., CT scans) to penetrate ancient artifacts and extract writing from manuscripts without disturbing them present both enormous opportunity and capacious storage requirements. Organizing and managing such diverse sources of information in a manner that supports the interests of users stresses digital library technology, in general, not to mention current approaches to metadata generation, in particular.

Larsen Paper submitted for the Digital Library for the Maghreb Workshop
Held in Rabat, Morocco, January 25-27, 2007

Access

Artifacts and manuscripts from the Maghreb reside in public and private collections around the world. Scholars conducting serious research into these materials traditionally incur large expense (both in time and money) to gain access. The potential of a MDLESC to provide uniform access to digital renderings of materials of scholarly interest (without endangering the materials themselves or imposing on their owners) is of immense scholarly value.

But lest we oversimplify through glib usage of terms such as “uniform access,” let us remember that these materials are very diverse and complex in terms of their representation. The metadata requirements are immense, and immensely valuable. As observed in the Chatham report, “metadata that incorporate translation, interpretation, analysis, and criticism – the digital library equivalents of the books and articles written about primary sources in traditional libraries – enhance and extend the use of material.”

Usage

Collections in a MDLESC offer unlimited opportunities for the development, testing, and utilization of tools to leverage the cognitive capacity of a highly distributed scholarly community. Collaboration tools that support spontaneous online meetings, for example, seem of particular value. Other “cognition-leveraging tools” identified in the Chatham report include collaborative editing, bibliography sharing, curriculum architecting, semantic tagging, knowledge mapping, visualization sharing, data-set structuring, and creating (shared or personal) logs or diaries of experiments and studies. Tools such as these, organized to support a global scholarly enterprise around a MDLESC, hold the promise of ensuring that the MDLESC will be a dynamic platform for discourse and discovery rather than a passive repository of static materials.

Stewardship

One major motivation for a MDLESC is concern over stewardship of ancient artifacts and manuscripts that may be at risk of being lost or destroyed. The promise of the MDLESC is that the collection and management of digital renderings of these materials not only increases their accessibility and (hence) value to scholars, but also provides a catalog to the original materials and a surrogate should the originals ever be lost.

But digital preservation is known to present its own set of unsolved problems, including the lack of proven storage technologies with substantial longevity and the rapid obsolescence of hardware and software. Add to this the direct threats from viruses and other malicious attacks, and it is clear that a MDLESC is not without its own challenges to stewardship. This suggests that a robust MDLESC may be centered in a major city such as Rabat, but will likely include a globally distributed set of repositories in order to provide resiliency over time.

Management

A large scale project like the MDLESC provides an unusual opportunity and necessity to consider organizational design. The MDLESC will likely develop into a global

Larsen Paper submitted for the Digital Library for the Maghreb Workshop
Held in Rabat, Morocco, January 25-27, 2007

consortium of participating organizations of varying shapes, forms, and expectations. It provides a potentially rich environment to foster new types of relationships, considering such factors as mission, governance styles, leadership, size, funding models, and an array of potentially challenging policy issues, including intellectual property rights, risk management, privacy, and security.

In summary, a MDLESC has the potential to become a major opportunity for scalable, interoperable digital library infrastructure serving a broad and diverse community around a thematic collection. An international collaborative effort of research, development, operation, and management can serve not only a MDLESC but also as an example of truly international collaboration serving the best interests of humanity.