

Challenges for Developing a Global Digital Library and Gateway: From *Global Memory Net* to World Heritage Digital Library

Dr. Ching-chih Chen, Simmons College

With the exciting convergence of content, technology, and global collaboration in this digital era, there are unprecedented potentials as well as challenges for developing digital libraries of all kinds. In the case of *Global Memory Net* (www.memorynet.org), its newly developed Interactive Multimedia Content Retrieval System (tentatively called i-M-C-S) has managed to incorporate several innovative concepts and functions to meet the challenges of a world digital library and gateway. This paper will articulate these challenges, sharing the experiences of *Global Memory Net*, and discuss briefly the exciting partnership with UNESCO's World Heritage Center in the development of World Heritage Digital Center (WHDC).

INTRODUCTION

Global digital library is not a new concept to me. For myself, as early as in 1993, I presented a keynote presentation at the 60th anniversary of the National Central Library in Taiwan [Chen, 1993] advocating a global digital library. But, in those early days, the concept was presented mostly from the “technological” point of view, contemplating the possibilities of linking the world digital collections together with the availability of a high-speed network like the Internet as we commonly knew that a couple of years later. In those early days, there was little idea regarding how to deal with the digital collections, and the Digital Library Initiatives of the US National Science Foundation (NSF) just begun. As time goes on, such a concept has moved rather rapidly to include R&D in many areas. Content has been considered to be just as important, if not more, as technology. In other words, with all the rich information resources in the world, if they are not digitally available, they will not be able to be accessed, retrieved, shared and enriched on the Web. Research activities related to content have proliferated in the areas of digital collection creation and development, organization and standards, interoperability and scalability, and many others. One important fact in relation to content is also becoming clear – no institution can provide everything. Thus, global collaboration in digital collection development becomes essential. In addition, it is natural to question how these digital resources are going to be used? The component of users and uses has surfaced as another important part of the digital library research.

This macro description, though oversimplified, is sufficient to provide a sketchy conceptual model, which coincides well with an interdisciplinary digital library research model, collectively developed by the members of the *DELOS-NSF Working Group in Digital Imagery for Significant Cultural and Historical Materials* [Chen et al, 2005] as shown in Figure 1. This figure shows clearly the triangular relationships among *people*, *content*, and *technology*. In other words, an interdisciplinary digital library needs to develop technologies to enhance the way all kind of users can use, share and contribute to the world digital content.

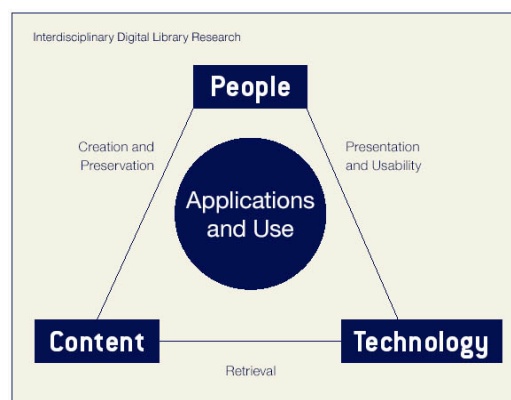


Fig. 1. Interdisciplinary Digital Library – Conceptual Model

Chen Paper submitted for the Digital Library for the Maghred Workshop

Held in Rabat, Morocco, January 25-27, 2007

CHALLENGES FOR DEVELOPING A WORLD DIGITAL LIBRARY

Conceptual model as described in Figure 1 is essential for developing a digital library because it provides a map to guide us in our planning and developing processes. Yet, one will face with many more difficult tasks in the process of actually creating an operational digital library application. This is because of the need to get down to the granularity of creating, developing and fitting the building blocks. Like completing a puzzle, all small and large pieces have to fit well together to create the complete whole.

In the case of *Global Memory Net (GMNet)*, since the beginning of our 5-year NSF/International Digital Library Project in 2000, we have conceptually set various requirements as our challenges for developing such a global digital library. The following are some of the selected ones:

- Using web as a platform to enable more user participation,
- seamlessly integrating multimedia information services to enrich user experience,
- providing innovative information services with the constructed databases,
- providing search and retrieval capabilities that can accommodate none pre-determined needs,
- enabling user control of the use of data, i.e. choosing either one single collection or multi-collections to search, language to use, etc.
- offering knowledge base and not just database(s), therefore once simple information is found, user can and should be enriched with more related multimedia information, and mixable or re-mixable resources if desired,
- providing freely accessible and searchable web resources, and not just links with them,
- offering multilingual capabilities for both presentation as well as retrieval,
- linking to the world collections as well as world bibliographic and web resources,
- addressing the desired and retrieved content in any granularity,
- enabling access of content geographical,
- protecting intellectual property of the content provider,
- encouraging the use of the discovered and retrieved data to develop own project(s) for whatever purposes, and
- enabling active contribution to the library by users.

Vision: As the name clearly suggests, *GMNet* also clarifies *my own vision of a Global Digital Library which is NOT a single digital library, but a network of multiple distributed ones, and that each can and will contain numerous high-quality multimedia digital content and materials.* Let me expand more in the following:

THE CASE OF GLOBAL MEMORY NET

Since 2000, *GMNet* has gone through many phases. As a small NSF supported R&D project, amazingly it has made considerable progress in many directions. Through the use of most appropriate cutting-edge technology, it has ensured the preservation of documentary heritage of world significance, and by public launching of its site (www.memorynet.org) in July 2006, it has made this heritage accessible to citizens of the world, both in and out of the countries in which it is physically located. Since then, we have witnessed numerous and on-going diversified use of this unique online knowledge base of the world cultural, historical and heritage from all types of users from all over the world.

Chen Paper submitted for the Digital Library for the Maghreb Workshop

Held in Rabat, Morocco, January 25-27, 2007

GMNet is an exciting collaborative and workable model at all levels for delivering multimedia content over the Web by utilizing cutting edge content-base image retrieval technologies in addition to the traditional metadata-based searching. It allows users to find images based on an integration of visual similarity and metadata relevance. The technologies used in the *GMNet* have been successfully tested on digital images of multiple cultures. While our current emphases are on culture and heritage contents, the i-M-C-S system can be used in various professional multicultural environments and subject areas.

- Content development – Currently *GMNet* is both a global digital library and gate. The fast growing number of collections from collaborators includes UNESCO’s Memory of the World, Asian Division of Library of Congress, national libraries of various countries, universities, museums, and archives, as well as our own core collection on the First Emperor of China. In addition to that, over 2400 world digital collections are also linked in such a way that they are searchable, thus the gateway function is not just a simple web-link. The infrastructure is ready to embrace more participation including digital contents for the Maghreb region.
- Technology development – Since 2004, we have developed our own in-house Linux/MySQL/PHP-based interactive Multimedia Content retrieval System (*i-M-C-S*) with much added functionalities to be elaborated in the following section. We also use SIMPLiCity for content-based image retrieval (CBIR) in collaboration with James Z. Wang of Penn State University.
- Global Collaboration – In addition to our international collaboration in the collection and technology development areas, extensive community building effort include the provision of R&D and education/training opportunities to 7 visiting researchers from China, Croatia, and Vietnam in the fall of 2005. These researchers have not only contributed to the content development by bringing relevant materials from their own countries, but also enhanced our technology capabilities with their knowledge and expertise in multilingual, web application, etc. All these have enhanced the capabilities of the i-M-C-S system. But, what is more important is their contribution to their own institutions and countries in terms of digital library development. For example, our visiting researcher from Croatia has become probably one of the most significant trainers of digital library application development in that country. He has also continued to be an important technical staff of *GMNet*.

Features of *Global Memory Net*

GMNet’s i-M-C-S system has many features and functionalities which are meeting the initial challenges described above. The following are some of the selected ones:

- *Instant access to rich image collections* – Using Web as a platform, one can access to over 30 image collections in *GMNet* with over 20,000 images as well as over 2400 digital collections from over 80 countries. Although *GMNet*’s own collections have focused in culture, history, and heritage, the 2400 digital collections in “World Digital Collection” cover all subject areas and include over 100 collections from UNESCO’s *Memory of the World*, over 290 collections from the US Library of Congress, as well as those from major national and academic libraries, archives, museums, etc. In other words, the world’s rich resources are instantly accessible. Figure 2 shows how these collections can be accessed.



Figure 2. Collection navigational panel

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

- *Seamless integration of multimedia information* – If relevant resource information on a chosen image is available in formats other than textual annotation, the user can then retrieve the relevant audio, video, etc. again by a single click of the mouse, as shown in Figure 7.

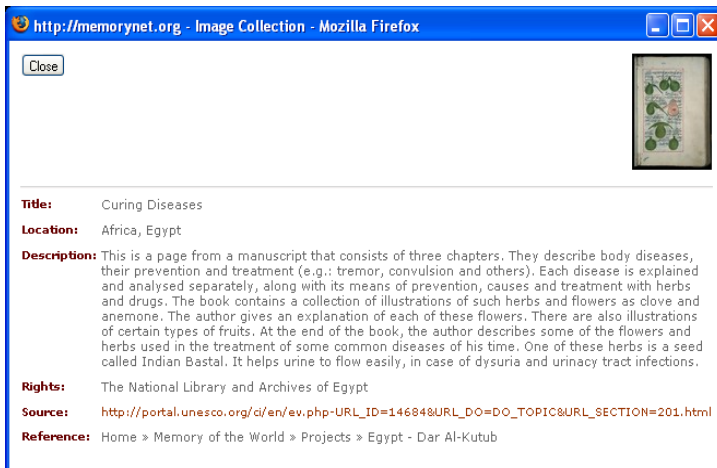


Figure 6. English descriptive information of the image. French description is also available.



Figure 7. The “Info” screen shows there are 3 digital videos and 2 PDF document files for instant retrieval. Some also has sounds and other file format

- *User can choose to conduct either a single collection or multi-collection search* – For example, if one wants to find information on “Africa” in both *GMNet* and the 2400 World Digital Collections, one can do multi-collection search. Figure 8 shows the search can locate instantly - 73 images in National Libraries, 63 Web sites in World Digital Collection, 159 images in UNESCO’s *Memory of the World*, and 7 images in World Musical Instruments.



Figure 8. Multi-collection search on “Africa”

We can also instantly locate the 64 world digital collections on the right panel. These include CultNet, Impression of Alexandria, Eternal Egypt, etc. All these can be linked instantly.

- *User can have much more in-depth learning of the subject by using expanded bibliographical and web resources* – For example, one can instantly find relevant books related to “Timbuktu”

Chen Paper submitted for the Digital Library for the Maghreb Workshop

Held in Rabat, Morocco, January 25-27, 2007

from the half-billion bibliographical records of OCLC/ World Cat, or web resources like Google, Wikipedia, Internet Archive, Million Books etc. (Fig. 9). This is how results of massive Digitization project are useful.

- *Multilingual display of descriptive information as well as multilingual retrieval* – Currently *GMNet* already has contents in English, Chinese, Croatian, French, Italian, Japanese, Spanish, Thai, Vietnamese, etc.
- *User can conduct geographical searches by continent and/or country* – Information on over 230 countries can be searched instantly provided the digital resources are available.

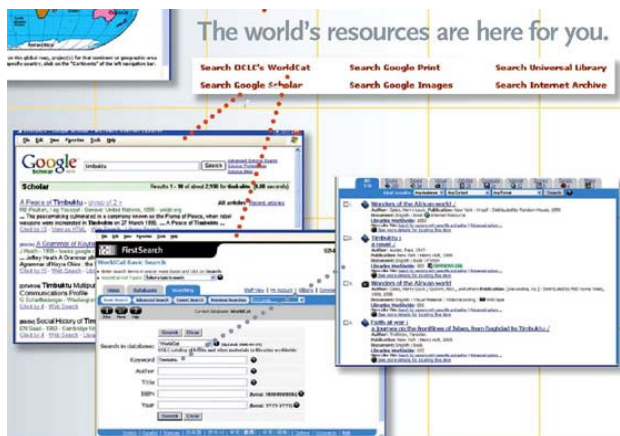


Figure 9. World resources are fully integrated.

In addition to these features, it is worthy to stress that *GMNet* positions its user to:

- participate actively during his/her own information seeking process,
- decide how he/she would like to use the discovered or retrieved information,
- develop his/her own project(s) by choosing, mixing and remixing the retrieved images, and
- be an active content contributor to *GMNet*.

WORLD HERITAGE DIGITAL CENTER

At the end of October 2006, a Memorandum of Understanding was signed between the UNESCO's World Heritage Center and Simmons College for the establishment of the first US World Heritage Digital Center (WHDC). The immediate primary activity of WHDC is to extend the use of *GMNet*'s technology to link all 830 world heritage sites from 138 countries together. Instead of the traditional access to individual images, users will be able to appreciate cultural and linguistic diversity through the search of visual similarity of images for cross cultural studies. This is an exciting development.

The UNESCO's List of World Heritage Sites shows clearly that many countries with rich and treasured heritage tradition and sites are also those where the digital library development is at its infancy. Many countries of the Maghreb region fall in this category. Yet, the cultural heritage is so rich here. For example, in Morocco alone, it has 8 world heritage sites. These include:

- Medina of Fez (1981)
- Medina of Marrakesh (1985)
- Ksar of Ait-Ben-Haddou (987)
- Historic City of Meknes (1996)
- Archaeological Site of Volubilis (1997)
- Medina of Tetouan (Formerly known as Titawin) (1997)
- Medina of Essaouira (formerly Mogador) (2001)
- Portuguese City of Mazagan (El Jadida) (2004)

In addition to the development of a world heritage digital library as an extension of *GMNet*, WHDC will involve in more R&D activities in areas like multilingual research, interoperability, etc. It will also promote the development of local heritage digital libraries in developing and less developed

Chen Paper submitted for the Digital Library for the Maghred Workshop

Held in Rabat, Morocco, January 25-27, 2007

countries by facilitating education and training of the heritage sites' digital managers and/or staff. Through more effort in capacity building, provision of knowledge in using open source software, such as Greenstone (currently Greenstone developer is *GMNet*'s partner), we can visualize the eventual success in digital library creation. Then, through OAI harvesting protocol, metadata of some of the useful local resources can be captured for the use of WHDC's application to benefit citizens of the world.

CONCLUSION

This workshop focusing on the Maghred region is not only timely but also badly needed. Let me end this paper by sharing some of the startling statistics of early 2006 from *GMNet* on the great disparity between the developed and developing world regarding the development of world digital collections. Not including the non-English ones in some regions(*), it is clear from the following listing that most collections have been developed by small number of countries in North America, Europe, Asia, Australia and New Zealand. Of the 230+ countries in the world, only one third of them have searchable digital collections. Of the 2347 world digital collections in *GMNet*, we found:

- North American (3 countries) accounts for 62%, and 56.3% of this is from the U.S.A.,
- Europe (50+ countries) accounts for 20.2%*,
- Asia (30+ countries) accounts for 12.3%*
- Australia and New Zealand accounts for 2%,
- Africa (50+ countries) accounts for 1.3%. Of this very small number, 1.1% represents digital applications that have been developed in Egypt and South Africa.

This statistics, though preliminary, is sufficient to show the current critical need not only to enhance the interest of many African entities with respect to the digital world, but also to find best ways to enable them to benefit from the existing rich global digital resources of all subjects, as well as to start creating their own local digital resources for preservation and universal access and use.

I truly welcome this opportunity to participate in this workshop for the creation of networked digital libraries of cultural heritage of the Maghred region, and to share our experience of *GMNet* and the latest development with WHDC. *GMNet* is a workable and tested model for global collaboration at all different levels/scope – big and small - in the development and delivery of digital multimedia content in culture and heritage using appropriate technologies. I look forward to discussing with any interested parties on how we can both contribute to and benefit from in this networked environment.

ACKNOWLEDGMENT

GMNet is supported by the National Science Foundation's International Digital Library Program under the Grant Nos. NSF/CISE/IIS-9905833 and NSF/CISE/IIS-0333036.

REFERENCES

[Chen, 1993]. Chen, Ching-chih, "Technological Potentials for the Global Digital Library," In: *Proceedings of the International Conference on National Libraries - Towards the 21st Century, Taipei, Taiwan, April 20-24, 1993*. Taipei, Taiwan: National Central Library. pp. 859-868.

[Chen et al, 2005]. Chen, Ching-chih, Howard Wactlar, James Z. Wang and Kevin Kiernan, "Digital Imagery for Significant Cultural and Historical Materials - An Emerging Research Field Bridging People, Culture, and Technologies," *International Journal on Digital Libraries*, 5: 275-285 (2005).