

Delivery of Web-based Multilingual Digital Collections and Services to Multicultural Populations: The Case of Global Memory Net*

Dr. Ching-chih Chen

Professor

Graduate School of Library and Information Science

Simmons College

300 The Fenway, Boston, MA 02115, USA

chen@simmons.edu

ABSTRACT:

This meeting focuses on addressing a number of questions such as how government libraries with multiple official languages or multiple languages that are in use in a country can:

- Plan and deliver services,
- Develop and acquire collections,
- Develop multi-lingual websites,
- Develop policies about service and collection development, and
- Other aspects of services to multi-lingual populations.

This paper will address some of the areas mentioned above by using *Global Memory Net (GMNet)* (<http://www.memorynet.org>), a multi-year International Digital Library Project supported by the US National Science Foundation and the forthcoming *World Heritage Memory Net (WHMNet)*, a partnership development with the UNESCO's World Heritage Center (WHC). *GMNet* enables one to access numerous digital multimedia and multi-lingual collections from many countries. *WHMNet* will provide access to multimedia and multi-lingual resources on 830 sites from 138 countries.

This paper will show that the cutting-edge system developed for *GMNet* and *WHMNet* will help government libraries deliver services to multi-lingual populations in a way not possible before, and should have great implications for governmental libraries. These include needed multi-lingual information resources and services to their multi-cultural citizens, as well as making multi-lingual resources digitally available for easy access.

INTRODUCTION

“Language plays a unique role in capturing the breadth of human diversity. We are constantly amazed by the variety of human thought, culture, society, and literature expressed in many thousands of languages around the world” (National Virtual Translation Center, 2006). NVTC provides interesting statistics on the 6,912 living languages by regions of the world in Table 1.

* Paper presented at the *IFLA Government Libraries Section* in cooperation with the *Library Services to Multicultural Populations Section* at the IFLA Meeting in Durban, South Africa, August 18-24, 2007.

Table 1. Breakdown of languages by world area

World area	Languages		Speakers	
	Number	Percent	Number	Percent
Africa	2,092	30.3%	675,887,158	11.8%
Americas	1,002	14.5%	47,559,381	0.8%
Asia	2,269	32.8%	3,489,897,147	61.0%
Europe	239	3.5%	1,504,393,183	26.3%
Pacific	1,310	19.0%	6,124,341	0.1%
<i>Totals</i>	<i>6,912</i>	<i>100.0%</i>	<i>5,723,861,210</i>	<i>100.%</i>

Yet, the five most spoken languages of the world by number of people are shown in the following with breakdowns of “1st language” and “2nd language”:

Languages	Languages with over 100 million speakers		
	Speakers (in millions)		Total
1st language	2nd language		
Mandarin	873	178	1,051
English	340	168	508
Hindi/Urdu	242	224	466
Arabic (all varieties)	206	246	452
Spanish	322	60	382

Given the large number of population of China and India, it is clear that English is considered as one of the most popular international languages. Yet, still, statistics has shown clearly that for most countries in the world, English is not commonly used. For example, French is likely to be much more popular in some countries in Africa, certain part of South East Asia, etc. Still, when English or French are used, they tend to be more by the more educated populations in many countries, and the local languages continue to be the predominant ones used by the general populations.

Also, many countries are multicultural, with some languages reflecting the earlier roots of settlers from other countries, like those from Europe, and some others representing those used by the newer settlers. In 2003, when Allison Dobbie of the Auckland City Council Library presented her paper at the IFLA meeting (Dobbie, 2003), she stated that her library provided at that time materials in 32 different languages in what they call the World Languages Collections. Some of these are historical reflecting their European roots and settlers from previous years, especially during the post war period. Some are new and developing rapidly in response to new settlement patterns. Chinese language collections were the largest and received the heaviest use then, with collection turnover of 8.2 per items per annual. She also listed that in her library, the languages which customers most wanted, apart from English, were Māori (the local ethnic minority group), various Pacific island languages, including Samoan, Tongan, Nuiean and Cook Islands; Mandarin (Chinese official language) and Cantonese; French and Spanish; Korean, Japanese and Hindi. It is fair for us to assume that we will find many different but sets of large number of languages wanted in other countries depending on geographical areas, migration patterns, political development, ethnic background, etc.

Thus, it is unquestionable that government libraries at different levels -- from federal to state or provincial, or to county, city and town – need to address the multicultural and multi-lingual needs of their citizens. In fact, in many countries, governmental published materials are required in both or multi-languages. For example, in Wales, Welsh and English are required by law for all their government publications.

While most libraries are pondering on how to develop and acquire multicultural and multi-lingual print collections, this paper concentrate in the development, organization, and access of multimedia and multi-lingual digital resources. In line with the general theme of IFLA 2007, “Libraries for the future: Progress, Development and Partnership,” this paper will focus on the use of cutting-edge technology to deliver multicultural and multi-lingual information services to the world citizens in a way not possible before. The system described on *GMNet* may seem to be a digital library for the future, but it is operational now, and it is available for universal access.

GLOBAL MEMORY NET (GMNet)

Global Memory Net (GMNet), a multi-year International Digital Library Project supported by the US National Science Foundation, launched its multimedia and multi-lingual web site for universal access in July 2006 (<http://www.memorynet.org>). It includes rich multimedia multi-lingual collections and resources from many countries. Its content collaborators cover all types of organizations and libraries, and even private individual with unique collections. Although two other papers delivered by this author at *IFLA 2007* are also relevant to participants of this section (Chen, 2007b; Chen, 2007c). They are of different emphases. This paper will focus on how multi-lingual and multimedia collections of governmental libraries and/or institutions at all levels are being organized in *GMNet* with the help of a cutting-edge integrated Multimedia Content Retrieval System (i-M-C-S) to enable the effective retrieval and delivery of multimedia and multi-lingual resources at a simple click of the mouse.

Sample Governmental Collections

GMNet's governmental content collaborators include organizations at many different levels – from international to national institutions, and from governmental libraries to museums, and archives. The multicultural and multi-lingual collections are numerous, and they include:

- UNESCO's *Memory of the World*, which include over 660 images selected from all projects of the UNESCO's Memory of the World, covering sample images of treasured collections from over 100 countries.
- US Library of Congress' Asian Division's unique collections include:
 - The LC's *Naxi Collection* – Naxi is one of the 56 ethnic groups of China. The collection of 3,342 Naxi manuscripts features ceremonial writings of the Naxi people of Yunnan Province, China.
 - The LC's *Inoh Map* - These are the first modern maps of Japan created originally by Tadataka Inoh between 1800 and 1821, and only recently discovered in the spring of 2001.
- *Emperor Collection* - The Emperor Collection serves as a model for integrating multi-format resources from national museums, archives, libraries and personal resources. The

original over 3000 images and related digital videos were supplied by the Ministry of Culture of China, the Bureau of Archaeological Relics of the Shaanxi Provincial government and the Museum of Qin Terracotta Warriors and Horses in Xian, the ancient capital of China.

- *Arab and Islamic Civilization* – An invaluable collection from the holdings of the National Library of Egypt, provided by UNESCO.
- *Thai Memory* - This unique collection consists of rare pictures of H.R.H. Princess Maha Chakri Sirindhorn from birth to the present, as well as those highlighting some of her recent memorable events and activities. These invaluable pictures are provided by the Thai Library Association and the Palace Archive of Thailand.
- *Taiwan Memory* – This include the invaluable old photos with descriptive information from the National Central Library (NCL) in Taiwan.
- *Vietnam Museum of Ethnology's* ethnic photo collections of the 54 ethnic groups in Vietnam. VME is the national museum devoted to preserve and present the cultural heritages of these ethnic groups.
- *Old Saigon* – This include the 350-year history of Saigon currently called Ho-Chi-Minh City. The photo collection is supplied by the General Science Library of Ho-chi-Minh City, which serves as the National Library of Vietnam in the Southern part of the country.
- *Maria Prymachenko Collection* – A collection of a famous Ukrainian folk artist, whose art works currently are scattered all over the country as well as in the Museum of Ukrainian Art.

From this sample listing, it is clear the digital multimedia collections came from many sources and countries in the world, thus multilingual functionality to access them is essential.

Selective Features of *GMNet's* i-M-C-S Retrieval System

GMNet is an exciting collaborative and workable model at all levels for delivering multimedia and multi-lingual 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.

If a collection is known to a user, he/she can conduct easy and flexible traditional searches by using any of the metadata fields known. However, this generally is not the case. We shall elaborate a few specific features taken from (Chen, 2007a):

- *User can explore the unknown collection and learn its coverage* – For an unknown collection of curiosity to the user, a government librarian can browse the image collection, but more significantly, the “random” feature permits the librarian to explore and learn the coverage of this collection. In seconds, the government librarian can learn how to retrieve effectively by the variety of images displayed as well as the words showing up for the titles as suggestive keywords. Figure 1 shows the diversity of images covered in the UNESCO’s *Memory of the World* collections for which a general citizen can try to locate desired images. In the case of

government librarian, for example, if the collected images are related to essential facilities and buildings in the country, the government librarian should be able to retrieve the images of the desired facilities in the same way.

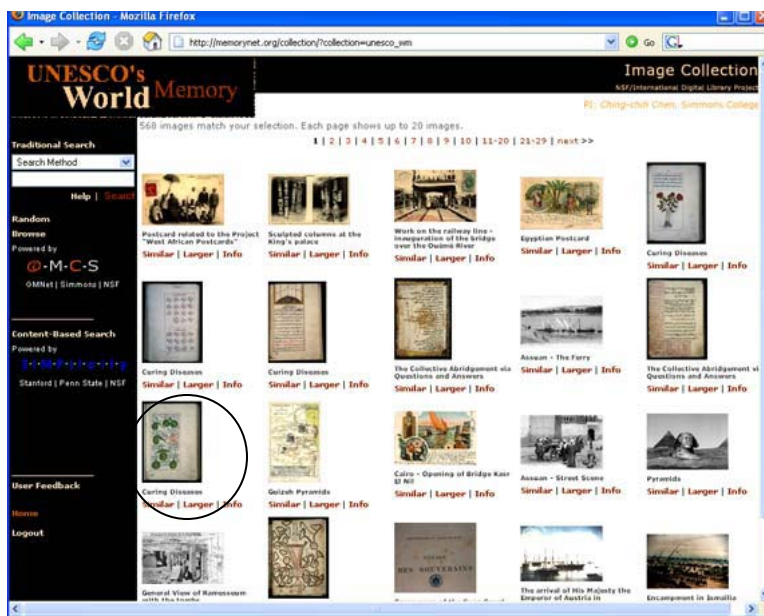


Figure 1. For UNESCO's World Memory, "Random" feature provides an instant display of a great diversity of topics covered

choices - to find pictures of the same color or shape using the cutting-edge content-based image retrieval (CBIR) technique (clicking on "Similar" – Fig. 3), or to zoom the image for larger sizes and more details ("Larger") (Fig. 3), or to obtain descriptive information ("Info") (Fig. 4).

- *Instant retrieval of similar images of interest by using cutting-edge retrieval techniques* – Once a picture of interest is chosen, for example, like the one on "curing diseases" as shown on in Fig. 1 and enlarged in Fig. 2, the government librarian is given three



Figure 2. Chosen Islamic image

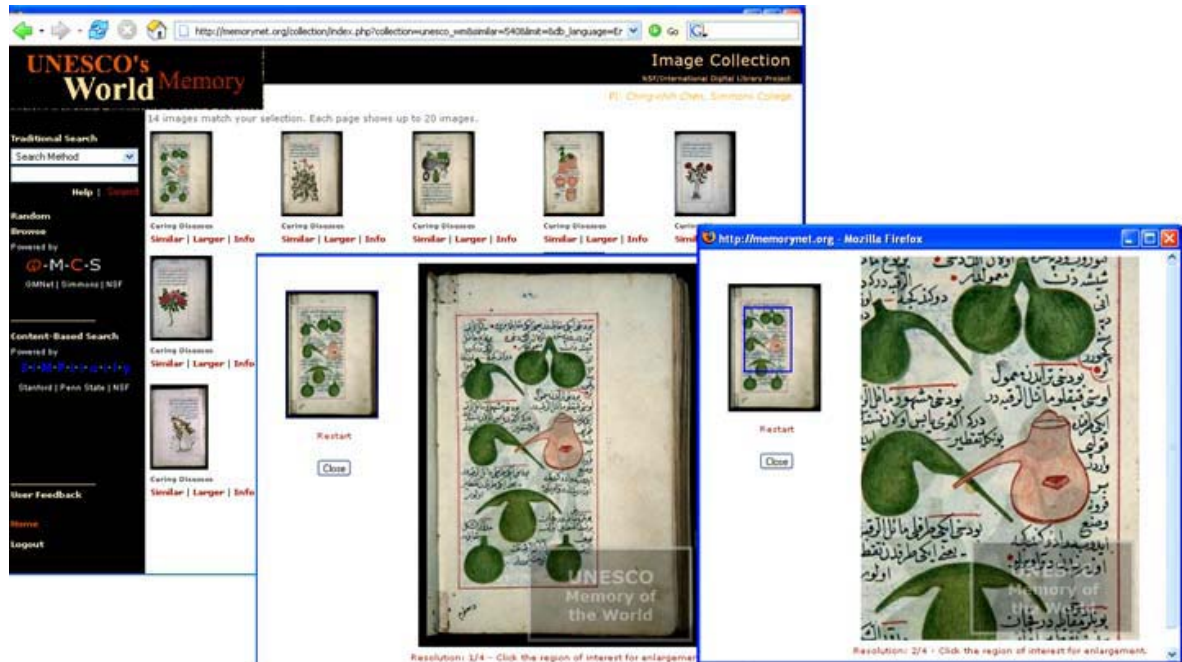


Figure 3. Instant display of all similar images of the same color and shape. A chosen image can be instantly enlarged to read the details of the Arabic text and digital watermark will be dynamically generated.

- *Dynamically generated digital watermark for copyright protection of the content provider* – Fig. 3 shows “UNESCO Memory of the World” as having the rights of this Islamic image (the digital watermark is in the lower right corner).
- *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 5.



Figure 4. English descriptive information of the Egyptian Image. French description is also available.



Figures 2-4 relate to images and materials found in a treasured rare manuscript at the National Library and Archives of Egypt. However, if one replaces this with any images in a government document available in multi-lingual format, one can imagine how readily how useful this capability would be to a government librarian in providing multimedia and multi-lingual information services.

← Figure 5. 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. For this one, the Chinese annotation is also available.

Multi-lingual Information Presentation and Retrieval

Presently, *GMNet* is able to present and retrieve about a dozen languages, including English, Croatian, French, German, Italian, Spanish, Russian, Ukraine, Chinese, Japanese, Thai, and Vietnamese, as shown in Figure 6. In fact, all languages in UNICODE are possible, thus we expect that the number of languages will increase quickly as the collections grow. This certainly has enhanced greatly the services to multicultural users, and has enhanced the usability of invaluable digital resources.

It is important to mention that all these multi-lingual texts are also retrieved as shown in the “traditional search” panel by simply typing in the desired term and selecting the specific field name or all fields of the metadata, as well as select the appropriate language.

In addition, all multimedia resources found are also instantly linked to available bibliographical print resources in all languages such as the half billion books in the OCLC database, as well as all the web resources, such as Google Scholar, Google Book Search, Internet Archive, Million Books, etc. All these contain resources in multi-lingual.

It is not difficult to imagine how a government librarian can greatly benefit from this capability in providing multi-lingual information to his/her users if descriptive information of graphic materials in government documents like land plots, county maps, etc. instantly.

A few more examples are given in the following to demonstrate additional aspects of multi-lingual services:

1. An image of the Library of Congress' Asian Division's Inoh Map in both English and Japanese (Figure 7). These invaluable Japanese maps are ancient government documents which are kept by the US Library of Congress, and not available in Japan. While physically they are not easily accessible, they can be freely and digitally accessed, retrieved and browsed with both English and Japanese descriptions on *GMNet*.



Figure 6. Annotations are available in over a dozen languages

2. An image of Maria Prymachenko's folk art called "Elephant." All records of this collection are available in 5 different languages – English, French, German, Russian, and Ukraine (as shown in Fig. 8 with the description).



Figure 7. Metadata information of the Inoh Map is available in both English and Japanese



- One of Maria Pryimachenko's paintings - Elephant. (English)
- Une peinture par Maria Pryimachenko - L'éléphant. (French)
- Einer der Anstriche von Maria Pryimachenko - Elefant. (German)
- Одна из картин Марии Прыймаченко -Слон. (Russian)
- Один з малюнків Марії Приймаченко -Слон. (Ukraine)

Figure 8. Brief description in 5 languages

3. Multi-lingual information is also further supported by sound. This is shown in the example of the ancient Waka poetry of the collection from University of Tsurumi. Fig. 9 shows that an ancient Japanese Waka poem is not readable even by current educated Japanese, because the writing used is not in contemporary Japanese. In this case, not only the annotations will be in both English and Japanese, but the reading in Japanese will be provided as shown in the sound files of that figure. A simple click of the sound icon, the Japanese reading will come up.

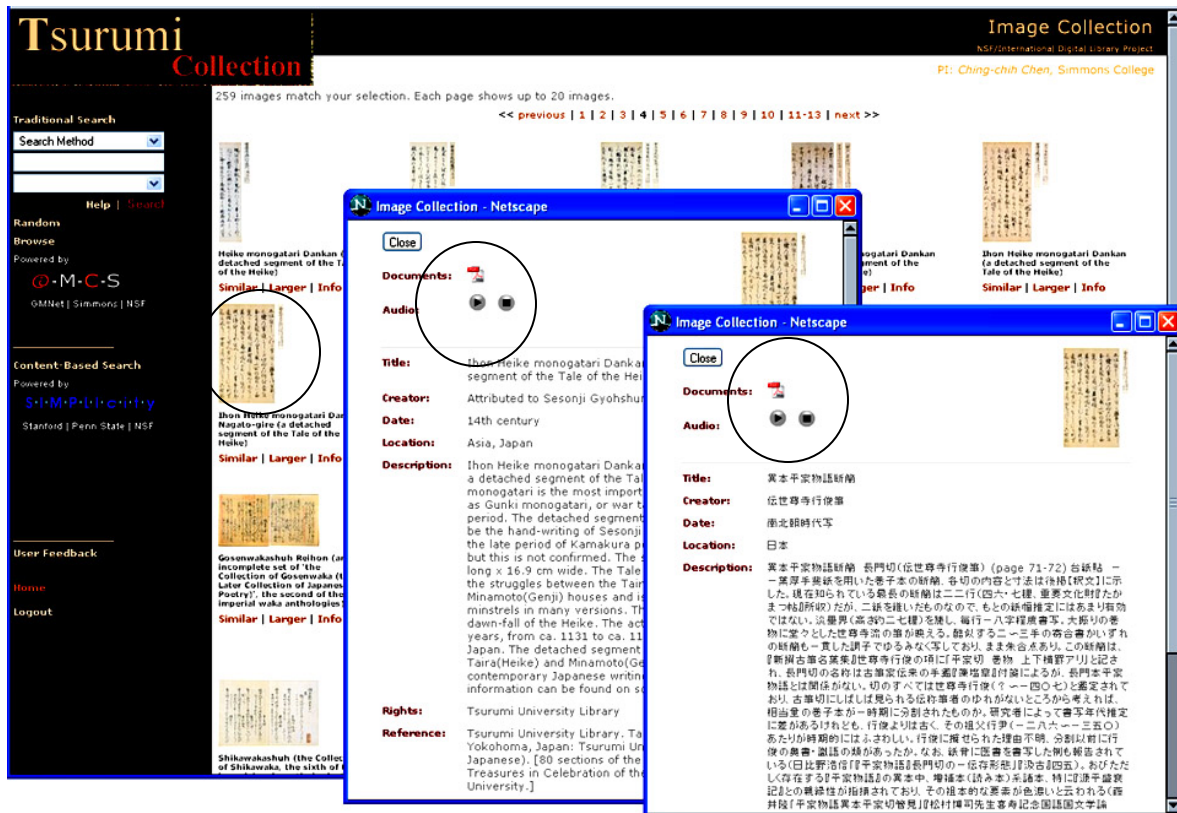


Figure 9. Ancient Japanese Waka poem can be described as well as read in both English and Japanese languages

WORLD HERITAGE MEMORY NET (WHMNet)

Another recent exciting development is our partnership with the World Heritage Center (WHC) of UNESCO – the establishment of the first US World Heritage Digital Center, which will use our *Global Memory Net* technology to provide multimedia and multi-lingual information on the 830 world heritage sites of 138 countries. See the announcement made by UNESCO/WHC in both English and French:

- English: <http://whc.unesco.org/en/news/326>
- French: <http://whc.unesco.org/fr/actualites/326/>

This is an exciting project with great challenges. From multi-lingual point of views, it is clear that 138 countries will likely cover great majority of languages in the world, if not all.

Like *GMNet*, *WHMNet* will start with images, with supporting annotations in multi-lingual format. The challenges in the multi-lingual areas for this project are many. They include:

- The uncertainly of all languages are available in UNICODE,
- The lack of language specialists as well as subject specialists on these WHC sites,
- The more dynamic need to search in one language but providing results in all available languages,

Clearly this will call for more aggressive collaboration among all stakeholders, as well as more cutting-edge R&D work.

CONCLUSION

GMNet and the forthcoming *WHMNet* are neither simple and traditional databases nor Web pages. They are really dynamic interlocking knowledge bases of the world cultural, historical and heritage digital resources. Numerous aspects of the systems will be described in other papers (Chen, 2007b, Chen, 2007c) to be presented at other divisional meeting of this IFLA conference. This paper focuses on *GMNet*'s ability to present and retrieve multi-lingual resources for the world's multicultural citizens. It should demonstrate that *GMNet* is an effective and innovative in using the cutting edge technologies and the available Internet and Web to deliver multi-lingual information services to general public in a way not possible before. Although the collections included in *GMNet* are focused in the world's culture, heritage and history, but the system developed is subject insensitive, and can be utilized for any types of government or non-government libraries in non-cultural areas. Thus, the technologies developed have great potentials for all government libraries which need to serve their multicultural users with needed resources in multiple languages. Some examples have been given in this paper on how a government library can benefit from systems like *GMNet*. Every government library owns many photographical materials in different formats – drawings; illustrations like maps – geological, typographical, road, etc.; and photos of ancient and contemporary persons, objects, and events. There have been no efficient ways of providing fast and easy multi-lingual information services to these. *GMNet* provides an exciting possibility for the innovative information delivery of this type of image contents as well as integrating information services with other types of resources.

ACKNOWLEDGMENT

GMNet acknowledges the invaluable support from the National Science Foundation's International Digital Library Program under the Grant Nos. NSF/CISE/IIS-9905833 and NSF/CISE/IIS-0333036.

REFERENCES

[National Virtual Translation Center, 2006] National Virtual Translation Center. Languages of the World. 2006. <http://www.nvtc.gov/lotw/>

[Chen, 2007a] Chen, Ching-chih, "Challenges for Developing a Global Digital Library and Gateway: From *Global Memory Net* to World Heritage Digital Library," Invited paper presented at the *Digital Library for the Maghreb Workshop*, Held in Rabat, Morocco, January 25-27, 2007, organized by the Fulbright Academy of Science and Technology. <http://www.fulbrightacademy.org/page/86609/>

[Chen, 2007b] Chen, Ching-chih, "The role of art libraries and their partners in documenting and preserving the threatened cultures: The Case of *Global Memory Net* and the World Heritage Digital Library," to be presented at the *IFLA/Art Libraries Section*, August 2007, Durban, South Africa.

[Chen, 2007c] Chen, Ching-chih, “Using Cutting-edge Technology for Innovative Global Resource Sharing Services: The Cases of *Global Memory Net* and World Heritage Digital Memory,” to be presented at the *IFLA/Document Delivery and Resource Sharing Section*, August 2007, Durban, South Africa.

[Dobbie, 2003] Dobbie, Allison, “Being Bicultural and Multicultural – The Challenge For Auckland City Libraries.” Paper presented at *IFLA, Library Services to Multicultural Populations Section Satellite Meeting*, 11-13 August 2003, Utrecht, The Netherlands.