GLOBAL MEMORY NET:
Potential and Challenges for Archiving and Sharing Cultural and Heritage Resources *

Ching-chih Chen
Professor
Graduate School of Library and Information Science
Simmons College, Boston, MA 02115, USA
chen@simmons.edu

Abstract:

This paper will stress the importance of digital content and community building for archiving and sharing cultural and heritage resources globally. It will discuss both the process related to the building of a global digital library. Experience of Chinese Memory Net and Global Memory Net, a major International Digital Library Project of the US National Science Foundation will be shared. Finally, the paper will elaborate on the enormous potential of collaborative effort in building such a global network. Implications to countries like India, which is unusually rich for its cultural, historical and heritage resources, will be discussed.

1. INTRODUCTION

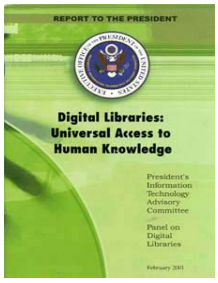

Figure 1. The *PITAC Report*

In February 2001, a *Report of the Panel of Digital Libraries of the US President's Information Technology Advisory Committee* entitled *Digital Libraries: Universal Access to Human Knowledge* (US PITAC 2001, Figure 1) was made public. The report clearly articulated that in the current networked digital environment, digital libraries must have enough contents in order to be able to provide universal access to human knowledge to any one, any where and any time. At the present, we are clearly a long way from achieving this vision.

In working toward this goal, it is truly impressive and exciting to learn that Dr. A. P. J. Abdul Kalam, President of India, launched a *Portal of Digital Library of India* on September 8, 2003. It is reported that this portal will have 27,000 or more books in digitized form available on the Internet to people globally by 2005. This is a part of the *Universal Library Project* of the Carnegie Mellon University (CMU) under the leadership of Prof. Raj Reddy. The project is otherwise known as *the India-US Million Book Digital Library Project* lead by Prof. Raj Reddy of the Carnegie Mellon University and Prof N. Balakrishnan of India. In addition to this gigantic effort, Profs. Reddy and I are also co-Principal Investigators of a sister project

* Presented as an invited keynote paper at the *ICDL 2004 (International Conference on Digital Libraries)* in New Delhi on February 26, 2004.

called the *China-US Million Book Digital Library Project,* which also involve in massive digitization activities in cooperation with the Chinese Ministry of Education and 6 major Chinese academic and research institutions for the first phase, and additional 6 other major universities all over China in Phase 2 (Figure 2). The first Phase institutions involved are some of the best known higher education institutions in China:

- Chinese Academy of Sciences
- Fudan University
- Nanking University
- Peking University
- Tsinghua University
- Zhejian University

Figure 2. Location map of the twelve major academic institutions with 2 regional centers (Courtesy of the *China-US Million Book DL Project*)

In addition to these India-US, and China-US collaborative efforts, more other big-scale activities are expected to follow. Just imagine, if we have half a dozen *Million Book Digital Library* projects in different part of the world, we will be able to build a substantial digital content base for universal sharing and use in a hurry.

In this paper, in line with the theme of this International Conference, and for what I am assigned to speak on, I would like to share with you the development and experience of another project supported by the International Digital Library Program of the US National Science Foundation (NSF), called *Global Memory Net (GMNet)*. *GMNet* has grown from *Chinese Memory Net (CMNet*): *US-Sino Collaborative Research Toward A Global Digital Library in Chinese Studies*. *GMNet* is meant to be a good compliment to the *Million* activities, because although the *Million* projects can and will include images and other multimedia contents, such as sounds, and videos, they are currently digitizing mostly those selected textual materials which are free of copyright concerns. Yet, *CMNet* and *GMNet* start with images and are currently working on videos and other multimedia formats.

For example, CMNet's core contents build upon the large quantity of visual materials of the author's earlier interactive videodisc project supported by the US National Endowment for the Humanities, entitled *PROJECT EMPEROR-I* with award winning interactive videodisc (1989) as well as multimedia CD (1991) products, both called *The First Emperor of China*. Figure 3 shows that it is related to one of the most significant historical period of China. The site is one of the most significant archaeological find of the 20th century. To each image included, extensive research efforts were made to provide relevant descriptive data (metadata) with annotations, as well as links to relevant references and texts whenever possible.

Figure 3. The First Emperor of China's Terracotta warriors in ancient capital Xian

CHINESE MEMORY NET (CMNet)

The NSF's supported *CMNet* is intended to develop a model for international collaboration with various R&D activities. It has made a great effort in developing collaborative

Fig. 4. The web-based *Chinese Memory Net*

infrastructure for digital library development with leading institutions in China, Taiwan and USA (Figure 4). For example, in addition to the development of extensive core Emperor image content and its association metadata at Simmons, *CMNet*'s collaborators in China and Taiwan are also engaging in various digital library activities, such as the Tsinghua University (Beijing, China) in Chinese architecture, Peking University (Beijing) in rubbings, and Shanghai Jiao Tong University (Shanghai) in Chinese music and musical instruments. To initiate collaboration, as a part of the collaborative research activities with *CMNet*, these three universities have developed a common OAI method to harvest the metadata of each institution for interoperable use of their cultural resources, and are currently testing the use of the extensive metadata of the Emperor images.

"In contemplating the future directions of digital library development, there is a cry for integrated functional libraries with large volume of digital contents in multimedia formats," and in distributed environment. Thus, the importance of content development and management was stressed in my paper on "designing multimedia digital libraries with content development and management" (Chen 2003) at the *DELOS/NSF Workshop on Multimedia Contents in Digital Libraries* in Crete, Greece, June 2003. Yet, the need to work together with the cutting-edge technologists in developing new tools and techniques in handling the large volume of multimedia contents is equally great. In this regard, the importance of cross-disciplinary collaboration has long been advocated by me, and illustrated clearly in one of my more recent books, *Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration* (Chen 2001, Chen 2001a, Figure 5).

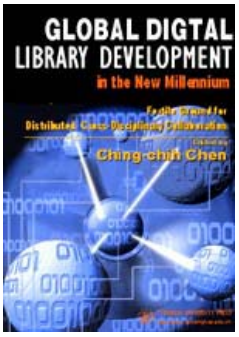

Fig. 5. Chen's book on cross-disciplinary collaboration

In the more sophisticated collaborative research and development area, *CMNet* has had some exciting results. Let me just list a few collaborative activities in the following for illustrative purposes (Chen 2003):

1. **Intelligent Agent and Image Retrieval** - Since 2000, the collaborative research with Prof. V. W. Soo of the Computer Science Department of the National Tsinghua University has yielded interesting results in developing similarity matching algorithm for retrieving relevant Emperor images (Figure 6). The preliminary Emperor's metadata have been used to develop ontological terms (Soo 2002; Soo 2003), and more extensive data will be used.

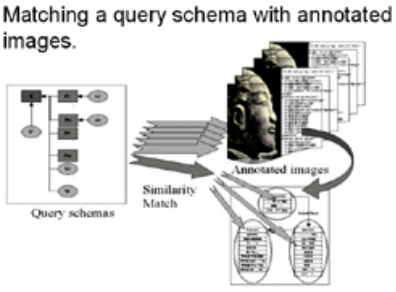

Fig. 6. Soo's work using Emperor metadata

2. **Semantic Sensitive Content-based Image Retrieval** - Since 2000, close collaboration with Prof. James Z. Wang of the School of Information Sciences and Technology and

Prof. Jia Li of the Department of Statistics of the Pennsylvania State University has enabled us to explore the use of semantic sensitive content-based image retrieval (CBIR) technology in web-based retrieval of *CMNe's* Emperor images (Figure 7) and other more advanced research (Chen 2002; Wang 2002; Wang 2003). Concurrently, the *CMNet*'s extensive metadata has provided invaluable baseline data for Wang's latest research supported by a NSF/ITR (Information Technology Research) grant. This is truly a good cross-fertilization example.

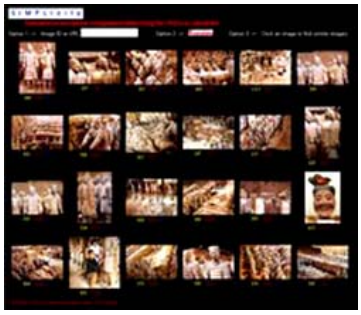

Fig. 7. Retrieval using SIMPLIcity

3. **Emperor's Digital Video and Informedia Technologies -** Carnegie Mellon University's well-known Informedia Project is one of the six original NSF/DLI-1 projects. It has continued its further development in digital video related technologies ever since 1995. Collaboration between Informedia and *CMNet* has enhanced perspectives from cultural and historical video documentaries (Figure 8). Its multi-lingual (English and Chinese) has also posed challenges in its speech recognition research (Wactlar 2002). When the Informedia technology is ready for web-based use, *CMNet* will be ready to explore it.

Fig. 8. Informedia's screen on the Great Wall

4. **Nanoparticle Technology in Restoration and Multimedia Digital Library Development -** Prof. Piero Baglioni and his research group at the Center for Colloid and Interface Science (CSGI) of the University of Florence have developed an effective and sophisticated nanoparticle technology for restoring damaged or degraded cultural relics. Currently, Baglioni's restoration images has formed the bases of our multimedia digital library, *Project Restore*, (Baglioni 2003, Figure 9), and has extended the *CMNet*'s scope and activities toward the *Global Memory Net (GMNet)*.

Fig. 9. Project Restore's image retrieval

GLOBAL MEMORY NET (GMNet) – ARCHIVING AND SHARING

As stated above, *CMNet* has established a model for international collaboration both for content development as well as for cross-disciplinary research. Once it is possible for developing a multimedia digital library in one subject disciplinary or for one geographical area, it is upward scalable to include more subject areas and bigger geographical areas. This was the

Fig. 10. The Home Page of *GMNet*

case with the expansion of the scope of *CMNet* to *GMNet* (Figure 10). The site map of the current *GMNet* will show clearly that the image and video contents of *GMNet* are currently organized and presented in the following hierarchy including the arrangement by:

- **Geography**
  - Continent - Asia, Europe, N. America, S. America, etc…
  - Country - Under each Continent, countries are listed alphabetically, such as China, India, Japan, Korea, Malaysia, Singapore, Vietnam, etc… under Asia.
  - Institutions or Projects - Under each country, multimedia digital collection can be arranged by institution, project, or combinations of both. This permits the quick retrieval of multimedia resources by a well-known collection, such as those from a given national museum.
- **Projects** - Special multimedia collections can be arranged alphabetically by project, such as *Chinese Memory Net, PROJECT EMPEROR-I, Project Restore*, etc. In addition to the hyperlink to the retrieval of each project's image collection, the extensive project archive and product(s) if any are also provided.
- **Archives** - This section provides all the archival materials related to each project.
- **Collaborative Activities** - This highlights all the major collaborative research related to *GMNet*.

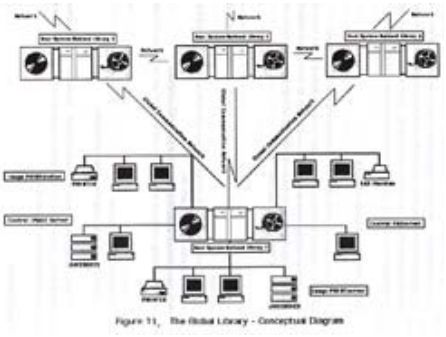

Figure 11, The Global Library - Conceptual Diagram

Realizing the coming of the powerful global information network, as early as in the 1990s, experience and knowledge gained from *PROJECT EMPEROR-I* prompted me to advocate digital knowledge base (Multi-media… 1994) with an early global digital library conceptual framework (Chen 1993, Figure 11). This over-simplified conceptual diagram suggests that if each country can develop its own large-scale contents digitally and properly organized them in interoperable ways, then in the current distributed networked environment, these digital collections can be linked together via the broad-band high-speed network from one country to the other to form a global digital library. This is what the *Million* projects are striving to achieve. It is in this same direction*, GMNet* is to provide a powerful portal which can link every worthwhile digital multimedia collection together. Technology is ready to do this, yet, in order to have a successful *GMNet,* technology alone is far from sufficient, it needs CONTENT! Without quality content, there is no digital or analog library. Yet, sufficient quantity of quality content can only be created with enough international communities which are ready to work together synergistically for this. This is why I am here to encourage the formation of such communities!

It is important to stress the great potential for pulling these contents together is enormous. There are many synergistic advantages, and I shall elaborate only a few implications for the participants of this international conference:

1. Although the word "memory" of the *Global Memory Net* is deliberately intended to cover all types of valuable memories, and can include many things. It can involve all subject areas, and of all time periods –historical or contemporary. Yet, currently, the multimedia content included in *GMNet* starts with those images and videos of cultural, historical, and heritage nature, as shown in the graphics above (Figures 3-4, 6-9). For this reason, it suggests that this portal should be of great interest and utility to the participants of this

conference. In addition, those with invaluable cultural, historical, and heritage contents should consider seriously as content contributors to this portal.

2. It is a win-win situation to contribute valuable multimedia content with appropriate metadata to the *GMNet.* In creating a digital multimedia archive, the contributing organization can achieve both purposes -- archival for preservation and resource sharing. In contributing some of the contents to the *GMNet*, it enhances the richness of the digital content of GMNet on one hand, and adds much global scope, greater breath, and more wide-spread utility to its own local collection.

3. Instead of duplicating effort in technology development, the management and retrieval tools and techniques developed and implemented for *GMNet* can be utilized to access the new collection. These include:

   - The effective but more conventional image retrieval method – retrieving by keywords, title etc., with zoom and other capabilities as shown in Figure 12.

   - Those cutting-edge content-based image retrieval techniques and tools as illustrated earlier (Figures 7 and 9). More real-time demonstration and discussion on this tropic will be given during the talk.

Fig.12. Emperor images retrieved by conventional methods

   - Dynamic digital watermark – knowing that intellectual property is a major barrier for digital library development, all contributors' image collection will be clearly protected with dynamically generated digital watermark indicating the ownership of the images.

   - Metadata creation and harvesting.

   - Video organization, management and retrieval – this will be worked on in the next phase of *GMNet*.

   - Etc.

CONCLUSION

Chinese Memory Net has created a model for global distributed collaboration in creating a global multimedia digital library. Recent effort has moved from the coverage of materials on Chinese studies to include those of any other types of areas studies. It has also moved from geographical coverage of China to covering any other parts of the world. Structurally, the *Global Memory Net* has been formed to accommodate more and more collaborative contribution. We are beginning to witness the realization of a small-scale global digital library with much richer and more diversified multimedia content to meet the needs of a much larger audience. *Project Restore* is one of such extended efforts.

In line with the recommendations of *the Report of the DELOS-NSF Working Group on Digital Imagery for Significant Cultural and Historical Materials* (Chen 2002a), we would

like to engage in more digital library community building, solicit more content providers to contribute their valuable resources, and increase collaborative activities in many essential fronts – content development and management on one hand, and cutting-edge technology development on the other. This is an exciting time for global digital library development. India has been in the forefront of developing a national digital library portal. With its history and incredibly rich cultural heritage, there are so much India can share with the rest of the world. For this reason, I hope that some of you will enthusiastically assert your leadership and join *GMNet* in our guest for a global multimedia digital library of cultural, historical, and heritage materials.

REFERENCES

(Baglioni 2003) Baglioni, P., R. Giorgi and C. C. Chen, "Nanoparticle technology saves cultural relics, and potential for a multimedia digital library," DELOS/NSF Workshop on Multimedia Contents in Digital Libraries, Crete, Greece, June 2-3, 2003.

(Chen 1993) Chen, C. C., "Technological potentials for the Global Digital Library," *Proceedings of the International Conference on National Libraries – Towards the 21st Century,* National Central Library, Taipei, Taiwan, April 20-24, 1993. Taipei, Taiwan: National Central Library, 1993. pp. 859-868.

(Chen 2001) Chen, C. C., ed. *Global Digital Library in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration.* Beijing, China: Tsinghua University Press, 2001.

(Chen 2001a) Chen, C. C., "*Chinese Memory Net (CMNet)*: A model for collaborative global digital library development," *Global Digital Library in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration.* C. C. Chen, ed. Beijing, China: Tsinghua University Press, 2001. pp. 21-32.

(Chen 2002) Chen, C. C. and J. Z. Wang, "Large-scale emperor digital library and semantics-sensitive region-based retrieval," Proceedings of the *Digital Library - IT Opportunities and Challenges in the New Millennium: An International Conference*, Beijing, July 9-11, 2002. pp. 454-462.

(Chen 2002a) Chen, C. C. and K. Kiernan, eds. *Report of the DELOS-NSF Working Group on Digital Imagery for Significant Cultural and Historical Materials*. December 2002. http://dli2.nsf.gov/internationalprojects/working_group_reports/digital_imagery.html

(Chen 2003) Chen, Ching-chih, "Designing multimedia digital libraries with content development and management: Technology potential and challenges," in *Online Proceedings of DELOS/NSF Workshop on Multimedia Contents in Digital Libraries*, Crete, Greece, June 2-3, 2003.

(Multimedia… 1994) "Multimedia and the First Emperor of China: Moving toward a digital knowledge base," *Multimedia Today*: 68-71(April 1994).

(Soo 2002) Soo, V. W., C. Y. Lee, C. C. Yeh, and C. C. Chen, "Using sharable ontology to retrieve historical images," accepted for presentation at the ACM/IEEE Joint Conference of Digital Libraries, Portland, OR, July 15-18, 2002. *Proceedings of the ACM/IEEE JCDL Conference, 2002.* pp. 197-198.

(Soo 2003) Soo, V. W., C. Y. Lee, C. C. Lin, S. L. Chen, and C. C. Chen, "Automated semantic annotation and retrieval based on sharable ontology and case-based learning techniques," full-length paper accepted for present at the ACM/IEEE Joint Conference of Digital Libraries, Houston, TX, May 29, 2003. *Proceedings of the ACM/IEEE JCDL Conference*, 2003.

(US PITAC 2001) President's Information Technology Advisory Committee. Panel on Digital Libraries. Report to the President. *Digital Libraries: Universal Access to Human Knowledge*. Washington, DC, February 2001.

(Wactlar 2002) Wactlar, H. D. and C. C. Chen, "Enhanced perspectives for historical and cultural documentaries using Informedia technologies," *Proceedings of the ACM/IEEE Joint Conference of Digital Libraries,* Portland, OR, July 15-18, *2002*. pp. 338 – 339.

(Wang 2002) Wang, J. Z., J. Li and C. C. Chen, "Interdisciplinary research to advance digital imagery indexing and retrieval technologies for Asian art and cultural heritages," *Proc. ACM Multimedia, Workshop on Multimedia Information Retrieval***,** Juan Les Pins, France, December 2002. 6 pp.

(Wang 2003) Wang, J. Z., J. Li and C. C. Chen, "Machine annotation for digital imagery of historical materials using the ALIP system," DELOS/NSF Workshop on Multimedia Contents in Digital Libraries, Crete, Greece, June 2-3, 2003.

About Prof. Ching-chih Chen

Dr. Ching-chih Chen is Professor of the Graduate School of Library and Information Science, Simmons College, Boston, USA. A sought-after international consultant and speaker in over 40 countries, she is an author and editor of over 35 books and more than 150 scholarly journal articles. She was the Chief Conference Organization of a series of 12 *International Conferences on New Information Technology (NIT)* from 1987-2001 in different parts of the world.

Currently she directs a major US National Science Foundation/International Digital Library Project, *Chinese Memory Net.* Using that as a model for archiving, content building in specifically image and video areas, organization and retrieval, and international collaboration, the project has expanded to "*Global Memory Net*" ready to include valuable cultural and historical resources of the world.

Served on the US President's Information Technology Advisory Committee from 1998-2002 with Prof. Raj Reddy of CMU, she is also the Co-PI of the *China-US Million Book Digital Library Project* with Prof. Raj Reddy. Active in the Digital Library area, she is the co-Chair of the *4th ACM/IEEE Joint Conference on Digital Libraries (JCDL)* of 2004 to be held in Tucson, Arizona in early June 2004. She will chair also the *NSF Workshop on International Digital Libraries* at JCDL. She is on the Advisory Board of DELOS (the European Digital Network of Excellence), US Co-Chair of the NSF/DELOS Working Group in Digital Imagery for Significant Cultural, Historical and Heritage Materials, and the Co-editor for the *Journal of Digital Library*'s Special issue on *Multimedia Contents in Digital Libraries*.