

# ***International Collaborative Digital Strategy: Promise for Libraries in A Knowledge-Based Society***

***Ching-chih Chen***

Professor

Graduate School of Library and Information Science

Simmons College

Boston, MA 02115, USA

Email: chen@simmons.edu

In the last decade, we have witnessed the incredibly fast-paced developments in digital telecommunications technology that has profoundly changed every aspect of our lives. In the Report of the President's Information Technology Advisory Committee's [PITAC] to the President, entitled *Information Technology Research: Investing in Our Future* [February 1999], ten critical "national challenge transformations" were identified:

1. the way we communicate
2. the way we deal with information
3. the way we learn
4. the practice of health care
5. the nature of commerce
6. the nature of work
7. how we design and build things
8. how we conduct research
9. our understanding of the environment
10. Government.

In just three short years, we have all personally experienced how our lives have been changed! Due to time constraint, I shall focus on how IT has changed "the way we deal with information." Today, most people in many countries have their options wide open in terms seeking and retrieving information. Although printed sources are still used by very large percentage of world populations, but to those countries with good Internet access, we have seen the dramatic change in information seeking behaviors from those of the past. With the powerful global network available to many of us at considerably low and affordable cost, more people have turned to the Internet and World Wide Web for all types of information.<sup>1</sup> Yes, the Internet has indeed changed the way we "deal with information" in very big ways, and have moved us up to the level that we are capable to truly talk, beyond "data" and "information," about "knowledge" in real substantive ways rather than using the word it as a fashionable "term." With the new trends, there are indeed major issues facing librarianship. If we know how to capitalize on the exciting and challenging opportunities ahead of us, we can have a spectacular role in this digital and networked knowledge society. If not, we will have the danger of being sidelined and ignored [Chen, 2000].

As I am talking about this, I cannot help to remember an early study of mine on citizen's information seeking. The study's results were reported as a baseline document for the first *White*

---

<sup>1</sup> While in 1994, there were a mere 3 million Internet and WWW users, it has grown to over 300 millions now, it will be more than one billion by the year 2005 (Chen, 1999).

*House Conference on Library and Information Services* in November 1979, and were the basis for the book, *Information Seeking*, by Chen and Herson [1982], which has been used widely for information need studies. During that early period more than two decades ago, it was shocking to find that of the citizens surveyed in six New England states, only 17% of them consulted library – in this case, we referred to public libraries -- in information seeking situations. Only 3% of the 17% mentioned library as the most “useful” information provider. That was the time when the Internet and WWW were nowhere to be seen by general public. I cannot help wondering how library is compared now with the Internet and WWW as viable and frequently used information sources?

At this celebrative occasion, I am not coming here to paint a “gloomy” future of the libraries. Instead, I would like to share with you the excitement of a global collaborative vision with tremendous potential in this digital, distributed, networked and knowledge-centered environment. I hope to explore together with you the exciting and changing time we are in, and how important it is for us to know what to do to prepare our libraries for this new era. If we can hold fast to our guiding principles on universal knowledge access and provision, and know how to use the available technologies effectively, we can make this time of change a moment of dazzling opportunity for all librarians in this knowledge-based society.

## **Digital Gift to the World**

Grossman and Minow in their book, *A Digital Gift to the Nation* [Grossman & Minow, 2001] quoted from the CEO of Hewlett Packard Corporation:

*I believe we are now entering the Renaissance phase of the Information Age, where creativity and ideas are the new currency, and inventions is a primary virtue, where technology truly has the power to transform lives, not just businesses, where technology can help us solve fundamental problems.*

Carly Fiorina, Chief Executive Officer  
Hewlett Packard Corporation

Then they went on to state in the Executive Summary of their book:

*In the age of information, the nation's prosperity, its democracy, its culture, and its future will depend as never before on the training, skills, ideas, and abilities of its citizens. The people's access to knowledge and learning across a lifetime in the sciences and humanities must become a national imperative in the emerging knowledge-based economy. In the past decade, information technology advanced the sciences and transformed the economy. In this decade, information technology will serve all of society and transform our daily lives.* [Grossman & Minow, 2001, p. 3.]

As we gather here celebrating the 50th Anniversary of this great library of the world, Shanghai Library, I doubt that I can articulate better the role of information technology (IT) in this emerging knowledge-based economy except adding the "global" concept. In other words, instead of articulating a nation's prosperity, its democracy, its culture, and its future, we need to expand to those of the world. To borrow the book title of Grossman and Minow [2001], let's consider our libraries' "digital gift to the world."

In expanding on this, while leaving many essential topics related to libraries and knowledge management to my Chinese and international colleagues here to elaborate, I would like to focus my talk on a few key concepts -- "digital," "global," "universal access," "knowledge access," "collaboration," and "strategy," as already suggested in the title of my talk. Given the nature of our work, my emphases will automatically on "Global Digital Library," which I advocated as

early as in 1993 [Chen, 1993], and have promoted the concept forcefully ever since in my speaking and publications activities. This focus should also match well with the theme of this international forum, "*Knowledge Navigation and Library Services: International Cooperation*. Because "seeing is believing," I shall take the liberty to accompany my talk with a number of visual demonstrations, which I apologize in advance for not being able to reproduce in this printed copy. This is one practical example of the power of "being digital" as I am borrowing another phrase, the title of a famous book from, Nicholas Negroponte, *Being Digital*, in 1996. Despite the true values of the originals of all hardcopy materials, such as books, journals, manuscripts, rare books, etc.; and analog resources, like slides, audios videos, and films, nothing can be shared via the powerful global network unless they are digital resources. In 2002, I believe that the differences between digital and analog are clear to most librarians and informational professionals, but the challenges are many. These include:

- How to built digital collections? What and how to digitize?
- How to preserve a digital heritage?
- How to organize and manage digital resources?
- How to find funding resources to build large-scale digital resources?
- How to create efficient IT infrastructure to accommodate these digital resources?
- How to integrate high-tech R&D results to create interoperable digital systems utilizing these digital resources?
- Etc.

These are common problems and challenges to everyone. Since many national libraries are represented here, let me use our largest national library in the US, the Library of Congress (LC), as an example. Arising from LC's own "sense of its vulnerability and uncertainty at the dawn of the information age and attempts to respond closely to the institution's own sense of its mission, LC asked to have a study conducted to provide strategic advice concerning the information technology path that LC should traverse over the coming decade." As a result, a report entitled *LC21: A Digital Strategy for the Library of Congress* [US. Library of Congress, 2000] was published. In the Preface of this reported, it states:

*The Library of Congress (LC) is a living and vital library and at the same time an icon. It is easier to be a library than to be an icon, but it is no easy thing to be a library amid the turmoil at the digital revolution... Every technology is spoken of as one that can store or transmit or search "the entire Library of Congress" in square inches of disk space or minutes or seconds of processing time... Inevitably, reality falls short of what the icon seems to promise [p. ix].*

The blue-ribbon Committee found the Library indeed falls short in realizing the promises of this digital age, and made numerous recommendations in areas listed above. I would recommend the report to all of you, but will give just a flavor on what was commented and recommended:

- The Library should explore cooperative arrangements for distributed collections more aggressively,
- Digital preservation should be a global concern, and the Library should have a long-term plan and strategy,
- The Library should treat the development of a richer but more complex metadata environment as a strategic issue,
- In "the world beyond its walls" section, the report calls LC to functions much less in isolation from its clients and peers, and must have more visibility in the digital arena.

The report further states that "the current transition to digital content calls for extraordinary, unprecedented collaboration and coordination," and LC needs to be "more proactive in bringing together stakeholders as partners in digital publishing and digital library research and development"[p.153].

From the global arena, it is safe to assume that the findings on LC should hold true to all national libraries. Thus, the recommendations made for LC as well as the strategies presented in the Report are truly worthy of our serious consideration. The call for LC to plan and strategize for "extraordinary, unprecedented collaboration and coordination" both in the US and globally should be of great relevancy to all national as well as large libraries in the world.

## **Digital Library Development**

In the midst of current incredible technological revolution, clearly the trend is to demand for faster, wider global network for more accurate, reliable and secured transfer of large-scale, multi-formatted information. This means that many services and activities that were not possible to offer and carry out due to the lack of network speed and bandwidth can be effectively explored now. In the areas of library and information services, many fiction-like concepts become reachable, more realistic, or even operable now. For example, universal access to information from a borderless global digital library or virtual library with unlimited multimedia resources is a technical reality. This will, for sure, challenge our traditional thinking of collection development. The traditional idea of "owning" gives way to "sharing" because no library in the world is possible to have everything. When library resources are pulled together, their richness is unmatched.

Clearly, projects related to large-scale digital libraries, high-resolution image databases, networked multimedia knowledge base, intelligent indexing and retrieval, distant education and learning, and the like will have great and direct impact on the way we do our business in libraries and information centers.

It is impossible for me to cover all these topics; I shall elaborate on "digital library" more. More information can be found in *IT and Global Digital Library Development* [Chen, 1999] and *Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration* [Chen, 2001].

## **Knowledge Management and Digital Libraries**

On July 11-13, 2001, I was invited to give a keynote speech at the *First International Conference on Knowledge Management* in Graz, Austria. The topic which I was asked to address is "Digital libraries (DL) as important ingredient of knowledge management (KM)." As I was preparing my talk, I realized that DL is not just the important ingredient of KM. In fact, DL and KM activities cover many common areas. Some of them are:

- Information Assets - What and where are the information assets of our institutions (organizational memory), how to acquire knowledge, etc.
- Information Overload - How to overcome this problem in order for us to provide the most relevant information easily (through hyper-linking process, relevance filtering, etc.)
- Knowledge creation
- Knowledge discovery - Through data mining using clustering, segmentation and other techniques, text mining, information visualization etc.

- Knowledge flow - Through collaboration, computer mediated communication, portals (information gateway), etc.
- Knowledge retrieval - Metadata and organization of knowledge, multi-lingual retrieval, intelligent agent, content-based image retrieval, digital video summation and analysis, etc.

### **International Collaboration and Universal Access**

It is clear that to address the above-mentioned areas, multiple skills, knowledge and expertise will be needed. In addition to library and information science professionals, we also need close collaboration with computer scientists, communications and network experts, and subject specialists. This is why collaboration crossing subject boundaries is essential. While digital libraries offer great potentials and promises, there are many unsolved problems and issues. Digital library R&D activities in recent years, particularly those supported by the US National Science Foundation's DL-1 and DL-2 initiatives, have made substantial contribution to the fundamental knowledge of digital libraries, yet major problems identified by Lesk in 1997, such as interoperability, scalability, sustainability, standards and methods for organization of digital resources, retrieval of large-scales multimedia/multimodal/multilingual digital resources, copyright and intellectual properties, etc. continue to be hunting us [Lesk, 1997].

Because of the enormity of these unsolved problems and issues, today's digital libraries, regardless what types, are poor shadows of the analog libraries in nearly every intellectual respect.

With this as a background, we can understand the need for international collaboration in order to achieve universal access of information via global digital libraries. Only through international collaboration, we can eventually fulfill the vision of the PITAC/Digital Library Panel's Vision that "all citizens anywhere anytime can use any Internet-connected digital device to search all of human knowledge" [US. PITAC, 2001].

In addition to sharing digital information anywhere anytime with anyone, international collaboration will also share national skills, avoid duplication of efforts, and create more harmonious standards.

To this end, the US National Science Foundation introduced its International Digital Library Project (IDL) in January 1999. IDL is intended to contribute to the fundamental knowledge required to create information systems that can operate in multiple languages, formats, media, and social and organizational contexts. The program's R&D activities focus on:

- Interoperable technologies;
- Technology for intellectual property protection in a global marketplace, and
- Methods and standards for ensuring long-term interoperability among distributed and separately administered databases and knowledge bases.

It is hoped that cooperative research can help avoid duplication of effort, prevent the development of fragmented digital systems, and encourage productive interchange of knowledge and data around the world.

### **Two US-China Collaborative Projects**

Let me share with you two US-China collaborative digital libraries which I am currently involved in:

1. *CMNet (Chinese Memory Net): US-Sino Collaborative Research toward a Digital Library in Chinese Studies*

This was one of the very first NSF/IDL projects funded in early 2000 to work toward a digital library in Chinese Studies, which is interpreted in the broadest sense. As the Principal Investigator (PI) of this R&D project, my collaborators came from the US, Beijing, Shanghai and Taiwan. Our research teams are most interdisciplinary with computer scientists, librarians and subject specialist, such as the research teams at Tsinghua University and Peking University in Beijing, Shanghai Jiao Tong University.

Understandably when we are researching into the cross-country, multi-language subject areas, we will face many difficult problems and issues which need to be explored and researched collaboratively. Some of the *CMNet* collaborative research areas include:

- Interoperable multilingual information systems;
- cross-language retrieval systems;
- multiscriptural and multicultural interfaces;
- distributed digital libraries including sound, data, text, image, and video;
- metadata techniques and tools;
- large-scale digital content development;
- preservation and archiving of digital scholarly information, including technology and procedures for long-term information asset management;
- social aspects of digital libraries and cross-cultural context studies;
- use of digital libraries at all levels of instruction;
- economic and copyright issues: authentication, rights, and fair use; and
- electronic publishing and scholarly communication technology, including laboratories, online repositories, and new methods of organizing knowledge distribution.

With such an ambitious list, clearly it is impossible for any one institution or one group to tackle all the problems and issues. This is where the benefit of having the “international” and “collaborative” approach. Let me just provide a couple quick demonstrations:

▪ Large-scale content and image data-base building

Capitalizing on the enormous analog and digital resources of Chen’s earlier Emperor Project on the First Emperor of China and his terracotta warriors and horses [Chen, 1991 and other source materials], a large image database consisting of several thousands very high-resolution images (some are scanned at higher than 1200 dpi with compressed TIFF file larger than 50MB per image for some). A demonstration exhibit was mounted using the LUNA Insight software. This illustrates vividly the potential of a quality high-resolution image base for scholarly research, preservation, etc. Figure 1 shows a screen of this exhibit, where 10 full-size terracotta warriors can be retrieved upon search by various metadata elements, such as keyword, title, etc..., and any portion of it can be enlarged. Furthermore on the very same screen, contemporary artist’s colored drawings can be retrieved and displayed to show the differences of what those warriors look now and should look before. Detailed metadata on the image can also be displayed (on the upper right column). Clearly, it is undisputable that this kind of access to images is far more superior than what is currently possible in most of our library and information centers.

What is the real labor-intensive work is the content building on each of the images. Figure 2 shows the background database record for each of the image. Without these

annotated information, even the retrieved image is what the user wants, it will mean not much.

▪ Sample Value-added International Collaborative R&D activities

Benefiting from the rich large-scale baseline content building, the project international collaborators can begin to conduct additional value-added R&D. For example, although CMNet is a very small (from funding point of view) and new program, yet in two years, we have already begun to see significant collaborative research. For example:

- The collaboration with Prof. Soo's Group (computer science) at the National Tsinghua University in Hsinchu, Taiwan on ontology and intelligent agent is to be reported at the ACM/IEEE Joint Conference on Digital Libraries (JCDL) in Portland, OR, July 15-18, 2002 [Soo, Lee, Yang & Chen, 2002] (see Figure 3).
- The collaboration with Prof. Howard D. Wactlar's Group (computer science) at the Carnegie Mellon University using their Informedia technology to extend to extend the historical perspectives using the Emperor's multilingual materials is also to be reported at the same JCDL Conference in Portland [Wactlar & Chen, 2002] (see Figure 4).
- The collaboration with Prof. James Z. Wang (computer and information science) at the Penn State University in the area of semantic content-based image retrieval using the Emperor data has shown great promise for handling both Chinese as well as the world art treasures and other subject images [Chen and Wang, 2002]. Figure 5 shows an image from the Web (downloaded from <http://www.unc.edu/courses/hist033/>), upper-left corner image, is used as the query to find related images in the Emperor database using the Simplicity technology.

These are just beginning example to show the "international collaborative digital strategy" of one small-scale NSF/IDL project to benefit from synergistic efforts, to avoid duplication effort, and to maximize the R&D results. I don't need to expand further on the promises for national libraries to collaborate in this knowledge-based digital age.

2. *China-US Million Book Digital Library Project (hereafter referred to as the Million Project)*

Differing from the small-scale US-China international collaborative R&D project, CMNet, this Million Project is a very large-scale international digitization project, co-sponsored by the Ministry of Education (MOE) of China, US/NSF, as well as other private sources. The US Co-PIs are Professor Raj Reddy of Carnegie Mellon University and Ching-chih Chen, and the Chinese PI is Prof. Gao Wen of the Chinese Academy of Sciences. I shall quote a few paragraph from the Memorandum of Agreement and the News Release of the MOE of China:

*"The Digital Library with its large quantity of digital multimedia information enables one to search for needed information through the Internet. The digital library's source data and information may come from libraries, museums, archives, universities, government sectors, professional organizations, or any person in the world, and its vision is to **provide access to all human knowledge in digital forms anytime, anywhere, via the Internet.**"*

*"...the biggest current global unsolved challenge is how to place large quantity of resources online, say at least 1-million books with text and images," in order to thus making it global accessible."*

*"It is expected that after two or three years, the China-US Million Book Digital Library Project will be able to place a substantial portion of the one million books online for global access via the Internet. Both the Chinese and American collaborators will create a world-class*

*technical platform with effective and efficient service and management for this proposed Global Digital Library.”*

*“In China, around 10 primary universities and institutes, including Peking University, Tsinghua University, and the Graduate School of Chinese Academy of Sciences in China will be organized to participate in this Million Book Project, led by the 211 Office of the Ministry of Education. Each participating university-based Digital Library Center has agreed to take on the needed research and development tasks in developing high-end computer systems to meet the needs of the Million Books Project. More significantly, each agrees to choose and offer its special, unique, and valuable information resources - books, paintings, sculptures and cultural objects - for this digital project. Each participating center will agree to digitize these resources under the sponsorship and supervision of the 211 of the planning Committee of the Ministry of Education of China”*

Clearly, this US-China Million Project and the US-India Million Project are the most ambitious international effort in systematically creating several terabytes of digital contents of high quality for universal access as of today. This can only be achieved through the high-level international collaboration. Once these digital resources are available, they are truly “digital gifts to the world!”

## **Conclusion**

In recent years, we have passed several "information jumps" -- from speech to writing to printing, and now to digital sources available via wire and wireless communications. For the libraries in this knowledge-based digital age, what lies ahead is a real jump for us toward universal information access via the globally connected distributed information and knowledge systems. With efforts like the Million Projects, concept of the “global digital library,” which I have been advocating for over a decade, is technologically feasible now. Yet, continuing R&D efforts are needed to address the problems and issues identified earlier through collaborative for large-scale distributed digital contents [Chen, 2001].

Clearly with these developments, the role of libraries and librarians for the 21st century will unquestionably change substantially. Now, it is clear that the lack of proper technology is no longer an obstacle. But, technology is not the end in itself rather the means to an end. We should use this marvelous technology to enable us “to do more by doing less” [Dertouzos, 1997]. As the librarians speculate on their work in this new millennium, what we must do is to make sure we can develop a vision for our future in this Internet-enabled knowledge society, and define its role in facing a new frontier. But, the greater challenge for us is to understand the evolutionary path and the transitional strategies that will be necessary to guide us from where we are today to any Next Generation Internet-related vision of tomorrow [Institute for Information Studies, 1998].

*“What makes this point in human history so interesting is that we are all -- people, organizations, and governments -- struggling with the vast changes in our processes and procedures that the Internet is bringing about, and will continue to bring about. The challenge is not to predict the end point of the change -- namely, what the Internet-enabled future will be -- but rather to better understand what we must do today and tomorrow, no matter what the Internet becomes or brings about... What people and institutions can do to cope with the changes the Internet will bring, however it emerges”*

[Institute for Information Studies, 1998].



This is truly a big order!

In closing, let me offer again my sincere congratulation and best wishes to Shanghai Library for its 50 years of glorious achievements. While sitting in this state-of-the-art auditorium of this gorgeous library, let us reflect on what it was like 50 years ago. I have taken the liberty in showing a few screens of the images of Shanghai Library then and now in 1952 and 1996 by using the Emperor's Exhibit system (Figures 6-8) together with another demonstration on what Shanghai was in the early 1970s and now.



Figure 6. Shanghai Library – 1952 and now



Figure 7. Enlarged picture section of Figure 6



Figure 8. Shanghai Library (1952 - top 4 pictures on the left and 1996 – lower 4 on the left, with enlarged front view of the new library on the right)

## References

- Chen, Ching-chih. (1991). *The First Emperor of China*, Multimedia CD-ROM. Santa Monica, CA: Voyager Co. Distributed by LearnTech (<http://voyager.learntech.com>).
- Chen, Ching-chih. (1993). "Technological Potential for The Global Library: Realities and Challenges," *Proceedings of the International Conference on National Libraries - Towards the 21st Century*, April 20-24, 1993, Taipei. Taipei: National Central Library. pp. 859-867.
- Chen, Ching-chih. (1998). "Next Generation Internet and the Academic Library's Role in Knowledge Society," *Proceedings of the International Conferences on New Missions of Academic Libraries in the 21<sup>st</sup> Century*. Beijing: Peking University, 1998. pp. 2-9.
- Chen, Ching-chih. (1999). *IT and Global Digital Library Development*. Newton, MA, USA: MicroUse Information, 1999.
- Chen, Ching-chih, (1999a) "NII, GII, I2 and NGI, and IT\*2 Initiatives: Implications to the Digital Library Development in the US," *in IT and Global Digital Library Development*. Newton, MA, USA: MicroUse Information, 1999. pp. 49-64.
- Chen, Ching-chih and Peter Hernon. (1992). *Information Seeking*. New York: Neal Schuman, 1982. p. 52.
- Chen, Ching-chih. (2000). "Next Generation Internet: Implications for Libraries in Knowledge Society," Invited keynote speech at the CONSAL 2000, Singapore, 26-28 April 2000. Appearing in its *Proceedings*.
- Chen, Ching-chih, ed. (2001). *Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration*. Beijing: Tsinghua University Press.
- Chen, Ching-chih. (2001a). "Chinese Memory Net (CMNet): A Model for Collaborative Global Digital Library Development," in *Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration*. Beijing: Tsinghua University Press. pp. 21-32.
- Chen, Ching-chih. (2001b). "Different Cultures Meet: Lessons Learned in Global Digital Library Development," *Proceedings of 1st ACM/IEEE-CS Joint Conference on Digital Libraries*, Roanoke VA, June 24-28, 2001.
- Chen, Ching-chih & Wang, James Z. (2002). "Large-scale Emperor Digital Library and Semantics-sensitive Region-based Retrieval," to appear at the *Proceedings on Digital Library - IT Opportunities and Challenges in the New Millennium: An International Conference*, Beijing, July 9-11, 2002.

Dertouzos, Michael. *What Will Be. How the New World of Information Will Change Our Lives.* San Francisco, CA: Harper Edge, 1998.

Grossman, Lawrence K. and Minow, Newton N. (2001). *A Digital Gift to the Nation: Fulfilling the Promise of the Digital and Internet Age.* New York: The Century Foundation Press.

Lesk, Michael. *Practical Digital Libraries: Books, Bytes, and Bucks.* San Francisco, CA: Morgan Kaufmann Publishers, 1997.

Negroponte, Nicholas. (1996). *Being Digital.* MIT Press, 1996.

US. Library of Congress. Committee on An Information Technology Strategy. (2000). *LC21: A Digital Strategy for the Library of Congress.* Washington, DC: National Academy Press.

US. President's Information Technology Advisory (PITAC). (February 1999). *Information Technology Research: Investing in Our Future. President's Information Technology Advisory Committee's (PITAC) Report to the President.* Washington, DC: National Coordination Office of Information Technology, February 1999.

US. President's Information Technology Advisory Committee (PITAC). (2001). *Digital Libraries: Universal Access to Human Knowledge (PITAC's Panel on Digital Library Report to the President.)* Washington, DC: National Coordination Office of Information Technology Research. Also reprinted with permission in *Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration.* Edited by Ching-chih Chen. Beijing: Tsinghua University Press, 2001. pp. 557-586.

Wactlar, Howard D. & Chen, Ching-chih. (2002) "Enhanced Perspectives for Historical and Cultural Documentaries Using Informedia Technologies," To appear at the *Proceedings of the ACM/IEEE 2nd Joint Conference of Digital Libraries*, Portland, OR, July 15-18, 2002.