CHINESE MEMORY NET (CMNet): A Model for Collaborative Global Digital Library Development

Ching-chih Chen

Graduate School of Library and Information Science
Simmons College
Boston, MA 02115, USA
E-mail: chen@simmons.edu

Abstract: Chinese Memory Net (CMNet): US-Sino Collaborative Research Toward A Global Digital Library *in Chinese Studies* is one of the very first collaborative research projects, supported by the NSF's International Digital Library Program (IDLP), which announced its first awards in May 2000. *CMNet* hopes to create a global collaborative model for distributed cross-disciplinary digital library development in different parts of the world. Although "Chinese studies" is the selected focus, yet the definition is deliberately broad to include all Chinese related subject fields, thus the "Chinese Memory." The projects' R&D results are expected to be beneficial to other "studies" as well, and expandable to include all types of collaboration in different parts of the world. *CMNet* will also play a significant role to the ambitious *China-US One-Million Book Project*, expected to be announced at NIT 2001.

This paper will provide an overview of *CMNet*, describe some of the initial R&D activities and results, discuss areas for potential productive collaborative research and developments, and encourage further enthusiasm on this type of collaboration.

1. INTRODUCTION

In 1984, when the Internet was still unknown to most people, and CD-ROM was not introduced to libraries, because of the high cost, the Library of Congress was the only library in the US, with the help of companies like SONY, could have videodisc technology applications. Recognizing then the potentials of the interactive videodisc technologies, I embarked on a major cutting-edge technology project in 1984, PROJECT EMPEROR-I, supported by the US National Endowment for the Humanities. This project demonstrated that interactive videodisc technology could indeed help to overcome the barriers of time and distance by enabling educators, researchers, students, as well as the general public to gain multimedia access to The First Emperor of China's fascinating terracotta figures of warriors and horses. In early 1990s, the project effort was expanded to use cutting-edge multimedia technology.[1]

---

[1] The enormously rich multimedia information resources in both digital and analog formats of PROJECT EMPEROR-I on the First Emperor of China's terracotta warriors and horses were used to create the award-winning products in interactive videodiscs (1991) and multimedia CD-ROM (1994), entitled *The First Emperor of China,* both published by the Voyager Company. PROJECT EMPEROR-I was support by the US National Endowment for the Humanities. Both are currently disbuted by the LearnTech Company in New York (http://voyager.learntech.com).

The experiences and knowledge gained from these multimedia R&D efforts in the 1990s made me realized the need to rethink the model for information dissemination and use. While databases were the "hot" topics and still today to many, I advocated the need to move toward "knowledge base" in an early featured article in the 1994 issues of IBM 's *Multimedia Today* (Figure 1) (Chen, 1994). Even a quick look of the images, one finds amazingly similar looks of much of the discussions on image retrieval of today.

Figure 1. Using Emperor materials to push for the "digital knowledge base" concept *(Color)*

With the advent of communications and information technologies and the coming of the Internet and World Wide Web (WWW) particularly in the 1990s, my involvement in the EMPEROR project, particularly the later ones related to the "digital" imaging and digital videos has convinced me of the enormous opportunities lying ahead of us as informational professionals. Thus, starting from early 1990s, I have advocated not only the concept of the "Global Digital Library" (GDL) but also the need to integrate digitally distributed information resources around the world in order to implement this GDL. The effort of *CMNet,* proposed in 1999, is in line with this effort, hoping to provide an actual model for collaborative work toward the realization of this GDL.

Why the GDL development is important? I shall borrow the arguments made in the US President's Information Technology Advisory Committee (PITAC)'s *Panel on Digital Library: Report to the President,* entitled *Digital Libraries: Universal Access to Human Knowledge*, February 2001 (see Appendix 1, pp. 559-586). In PITAC's February 1999 report, *Information Technology Research: Investing in Our Future*, the Committee listed 10 national challenge transformations as "essential prerequisites to enabling all citizens to participate fully in our society and to benefit fully from the Information Age." These transformations are:

- The way we communicate
- The way we deal with information
- The way we learn
- The practice of health care
- The nature of commerce
- The nature of work
- How we design and build things
- How we conduct research
- Our understanding of the environment
- Government services and information

To PITAC, digital libraries will play a central role in all the transformations. Thus, the same argument can be forwarded to "global digital libraries" when all citizens of the world are considered.

2. *CMNet*: WHAT IT IS AND WHAT IT HOPES TO ACHIEVE?

2.1. Driving Forces of *CMNet*

With the growing interest in and need for education and research in academic institutions world wide, there is an urgent need to develop an effective collaborative worldwide system for sharing rich information resources. Currently most of the needed information is not available in digital form. While efforts in converting them to digital collections are increasingly obvious in various parts of the world, the activities are haphazard and are being introduced with insufficient concern for interoperable methods, standards and technologies. Furthermore, almost all efforts in many countries are not coordinated internationally. Thus, a pressing question would be:

> *How to develop effective international cooperative research that can help avoid duplication of effort, prevent the development of fragmented digital systems, and encourage productive interchange of essential knowledge and scholarly data around the world?*

To facilitate the much education and research, there is a great need for the creation of sustainable and well coordinated global subject digital systems that can operate in multiple languages, formats, media, social and organizational contexts. To achieve this in the foreseeable future, we need to engage in international collaborative research that can start to address many primary research questions. These include:

- What are the significant digital collections on a given subject available for research and education and how to locate them?
- What are the common methods and standards for formatting and organizing such digital collections for global interoperable retrieval, dissemination, and use?
- How do we develop a suitable technology for intellectual property protection in the development of linked, compatible databases?
- What multilingual problems will surface when accessing digital collections, specifically with the materials in non-roman languages?
- How to develop effective content-based multimedia retrieval system for accessing these digital resources?
- How can we access such globally-distributed information through currently available networks of varying speeds and bandwidth -- how can we overcome technical barriers, system fragmentation, or other limits?
- How do we evaluate the effect of this new technological approach from social, cultural and educational points-of-view?

Looking for answers to some of these questions have motivated me to propose *CMNet*. Choosing "Chinese Studies" other than other subjects has clearly been influenced by my earlier EMPEROR project and the significant and high-level connections that I have cultivated in the recent years. All NSF international projects requires matching grants from the participating countries.

2.2. *CMNet*'s Goals and Immediate Objectives

The long-term goals of *CMNet* are:

- to contribute to the fundamental knowledge required to develop and create an effective global information system for sharing Chinese cultural and heritage resources for academic

education and research, that can operate in multiple languages, formats, media, and social and organizational contexts, and

- to enhance academic users' ability to access and exploit significant digital collections in Chinese studies in new ways, regardless of location, language or formats, and enable broader use in research and education.

The project's immediate objectives are:

- to avoid duplication of efforts,
- to prevent the development of fragmented and non-interoperable digital systems in Chinese studies already taking place in different parts of the world,
- to define characteristics of an interoperable system in dealing with Chinese information resources,
- to find possible solutions for multilingual (specifically Chinese) language processing,
- to develop commonly accepted standards of methods, protocols, tools and technology for system development in the current networked environment,
- to foster sustainable relationships between Chinese and American academic and research institutions initially and then move forward,
- to explore, develop and establish workable international collaborative mechanisms for more future research efforts in expanded subject areas, and
- to connect to other major national and international collaborative digital library projects.

3. CMNet's Research Approaches and Activities

CMNet has first brought selective academic educational and research partners in the Beijing, Shanghai, Taipei, and the USA together working toward an effective and sustainable global digital library in Chinese Studies. The collaborators and affiliates of this complimentary and synergetic group possess experiences, knowledge, expertise, and capabilities in different but related research areas with their research activities funded by their respective governmental agencies. Several participating institutions will bring to the research some superb collections of Chinese culture and heritage, which form the current core of *CMNet*'s Chinese memory. As the collaboration grows, the content and scope related to the Chinese memory will grow accordingly so will the research activities and techniques involved. These collections are unique and essential to education and research, but currently are not accessible or usable via global networks because of distance, form, and technical barriers. This research project is to find new ways to enable academic users to access and exploit these significant research collections via global networks. The R&D activities will also contribute to enhance the capabilities and effectiveness of the collaborative digital library development.

3.1. Initial Collaborators and Their Expected R&D Activities

Currently, our Chinese collaborators include:

Beijing -- Peking University and Tsinghua University
Shanghai -- Shanghai Jiaotong University
Taipei -- Academia Sinica, Taiwan University, and Tsinghua University (Hsinchu)

- The US collaborators and project affiliates are still being formed. However, preliminary exploratory activities have already started. Several potential affiliates are either presenting papers at *NIT 2001* (such as Zeng, p. 405) or presenting preliminary *CMNet* related results (Qin, p. 243).
- The Taipei collaborators' funded research projects include the following:

- o   Agent-based information gathering and service for digital library (Soo, et al, p. 299)
- o   Intellectual property rights protection -- Safeguarding digital library contents and managing electronic copyright
- o   Linguistic technology and resources for English-Chinese bilingual information system (Chen, Hsin-His, p. 33)
- o   Lexicon-based knowledge linking -- Approaches towards a WordNet
- o   Metadata interchange for Chinese information for digital libraries infrastructure for multilingual digital library.

The Taipei collaborators will use EMPEROR data for cross-test on their R&D results, as exemplified in the research described in Soo, et al. (p. 299).

- The Beijing and Shanghai collaborators' proposals are still pending for decisions from the Natural Science Foundation of China. If funded, their research will involve in areas:

- o   Digitizing techniques for Chinese inform. – Include pattern analysis & recognition for ancient books; video/audio structuring process; interactive 3D display techniques; etc…
- o   Metadata techniques and standards for Chinese information (Xiao and Chen, p. 363)
- o   Intelligent HCI techniques – Include multi-lingual, multi-modal search and retrieval; translation engine; etc.
- o   Digital library architecture – Include interoperability & scalability study on Internet environment; management framework and organization schema for distributed, heterogeneous, and mass information repositories; etc.
- o   Theory/method on global Chinese information organization, integration, and sharing
- o   Prototyping systems for Chinese cultural and heritage information (such as Xing, et al, p. 373; Wang, S., p. 341).

Clearly, the partners will draw on their substantial experience in database construction, multilingual information system, metadata, digital imaging, network development, and the like to lay the groundwork for a new and effective collaborative information system. This system will transcend present inhibitors such as geography, time, and the cost of international consultation; and open up access to Chinese cultural and historical information resources and artifacts for a distributed set of international users taking advantage of the currently available advanced global network capabilities. This will undoubtedly broaden the use of and access to the unique digital collections in Chinese studies at a later date by knowledgeable scholars all over the world.

3.2. *CMNet*'s Research Activities at the Home Base

In addition to all the coordination and networking activities with the Chinese and American collaborators and affiliates, at this first stage, the Simmons research team, under my general direction and supervision, will capitalize the already existed and very rich analog and digital collection and database information of the earlier PROJECT EMPEROR-I to create a prototype multimedia system for exhibition and further R&D purposes (to be discuss in greater details in Section 4).   This prototype system will enable the project collaborators and affiliates to use a small segment of the large Emperor digital resources for their research as well as for testing purposes.   Some of the preliminary plans and results are already reported at this conference (Qin, p. 243; Soo, et al., p. 299).

3.3. *CMNet*'s Broad Research Areas

Since each collaborating institution mentioned above is at a different stage of "digital library" development and each research group is exploring different research topics, each has tentatively identified the division of research in this collaborative environment. In further

delineating the research areas, it becomes clear that the project will capitalize on the rare international research synergy to explore a number of research areas including:

- Multilingual information systems, and cross-language retrieval systems;
- Multimedia and multi-format digital libraries including sound, data, image, videos, software, and other kinds of content;
- Interoperability and scalability technology to permit large world-wide collections;
- Metadata techniques and tools;
- Linking Chinese historical, cultural and heritage information systems of some of the best institutions;
- 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;
- Utilization of digital libraries in educational technology at all levels of instruction;
- Economic and copyright issues: authentication, rights formalism, and fair use; and
- Electronic publishing and scholarly communication technology, including collaboratories, online repositories, and new methods of organizing scientific knowledge distribution.

4. *CMNet*'S MULTIMEDIA PROTOTYPE SYSTEM

In order to provide a prototype system capitalizing the existing resources and databases from the EMPEROR project, and also making it possible for our collaborators and affiliates to utilize the data for further research, we have chosen to utilize a commercially available software system for visual collections, Insight® of Luna Imaging, Inc. It is a software that, as claimed by Luna:

- Works with visual collections over the Internet,
- Uses images of virtually any resolution,
- Stores and distribute images in multiple resolutions and formats;
- Zooms, pans and arranges images, and
- Creates and shows presentations.

The system architecture of Insight™ is shown in Figure 2. The feature of the system, as claimed by Luna, is to "provide images in ways that are natural and conducive to thinking with a visual medium." Yet, it is still a keyword-based system, thus it has the inherent limitations of the keyword approach, as articulated in several papers in this book (Gao, et al., p. 71; Zhang, et al., p. 429).

4.1. Why Insight™?

One may wonder then why a R&D project such as *CMNet* is using a commercial software now, rather than developing its own system. Here are the reasons:

1) **Need to set up a prototype system quickly**. As stated earlier, in order for our own research team (including our affiliates) as well as our collaborators to assess the Emperor multimedia data for research potentials. A prototype system on the Web should be set up quickly. Although *The First Emperor of China* multimedia CD was created in 1994, yet it is still a popular CD now, which enjoys wide distribution. As a desktop multimedia product, users still enjoy and amaze at the CD's user-friendly interfaces and its rich and hyperlinked multimedia information resources. However, these were accomplished in the early 1990s.   In 2001, we should do much better than those accomplished in 1994. The current digital library system should provide the kind of functionalities, retrieval capabilities, multilingual and multimedia information to users in a way not possible before. A prototype will provide us a base to work for better. Considering the cost involved, there is no easier, better, and faster way to

accomplish this other than selecting one of the most suitable and capable software in order to meet our needs.
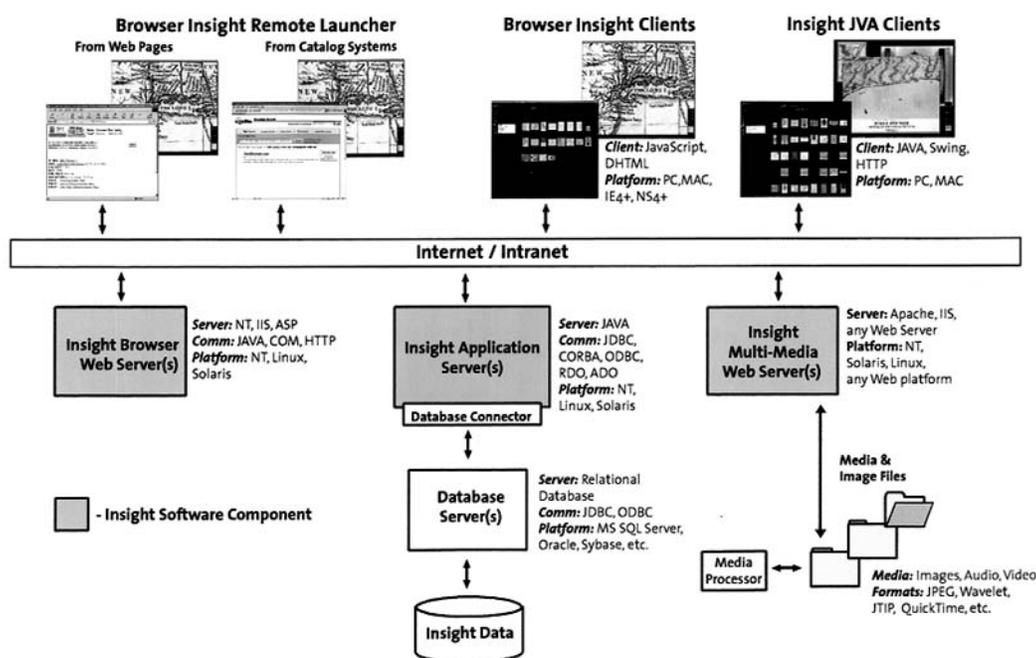
Figure 2. Insight™ system architecture *(Color)*

2) Our multimedia prototype exhibition collection is in good company with a number of reputable institutions whose digital image collections are also not developed in house but use the Insight™ system. These include Columbia University Libraries, Cornell University and its Herbert F. Johnson Museum of Art, Harvard University Libraries, Massachusetts Institute of Technology, Mellon Foundation, National Gallery of Art in Washington DC, The National Museum of the American Indian, New York Public Library, Princeton University, Standford University, University of California at Berkeley, UCLA Hammer Museum, and Yale University. This means that although many of these institutions are active in their digital library development activities, they also have the similar approach and idea as *CMNet*, e.g. to use the commercial system when that can meet their needs.

3) **Data compatibility with the Insight™ system**. *The First Emperor of China* multimedia CD was one of the earliest multimedia projects in the US. Both interactive videodisc and CD products received several industry awards and recognitions, including the American Association for Visual Communicators' Silver Cindy Award in 1991. The CD was chosen by *MacUser* as one of the 50 best multimedia CD-ROMs in the market place for the year of 1994. With 108,000 analog images for the popular version of the videodisc, it was an incredible challenge to provide interactive access to that large number of image frames. 1994 was a time when multimedia products were produced with no guidelines, and information organization for those multimedia information resources had no information organizational standards. Metadata was not even a concept then. Yet, images had to be organized, and effective retrieval methods have to be developed. Out of desperation, I developed my own way of organizing, cataloging, and indexing each essential videodisc images. Figure 3a shows the 1985 original record structure for the selective 1900+ images, while Figure 3b has been reformatted with current headings but still with the 1985 original data. Amazingly, these consist of a number of basic elements of the current popularly used metadata standards, such as the Dublin Core and the VRA. Even a quick look at these raw data, one can agree that this database consists of dynamite resources and MUST be fully utilized. They also represent enormous human intellectual efforts and cannot be matched by any machine generated indexing activities. For example, in 1985, realizing that the simple keyword

generated from title or the description (even it is a long one) cannot satisfy the retrieval demand for scholarly use.   Painful efforts were made to "content-index" each image.   For example, to the image of Figure 3a, it is not just a warrior's head; it is one with a side-view. It has specific kind of mustache, and brushed hairlines, etc. These are evident from the large number of keywords given to the image.   Thus, with this invaluable original raw data, every effort is made to fully utilize them for the Insight™ exhibition system.

4) **Metadata Cross-Walks** - Several papers in this book have discussed the desirability of metadata cross-walk schemes (Xiao and Chen, p. 363; Zeng, p. 405).   Insight™ system can present the Emperor data in several different standards -- CIMI, CD-WA, Dublin Core, VRA, and US-MARC, in addition to our own formats.   This is deemed to be attractive since creating data confirming to any of these standards and also creating cross-walk data are simply too time-consuming. However, further study is needed to determine its accuracy.

~Warrior's head - moustache
V=48769   B=48443
D=
J=c. 200 BC
M=Lintong
P=Pit #1, Lintong
R=Pit #1 Excavation Report - 1974-84. (Beijing: Wen Wu Publishing, 1988). Vol. 2, p. 117.
W=Black & White
K=Soldier,Soldier-head,Soldier-head-close-up,Soldier-terracotta,Warrior,Warrior- head, Warrior-
    close-up,Warrior-terracotta, Face,Face-side, Face-right,Face-close-up,Faceterracotta,Head,Head-
    close-up,Head-side,Head-terracotta, Moustache,Moustache-right, Moustache-side,Moustache-
    terracotta,Hair,Hair-side,Hair-right,Hair-close-up,Hair-terracotta,Ear,Ear-right,Ear-close-up,Ear-
    terracotta,Qin, Terracotta,Qin,
X= A thick moustache, the detail of the right ear, and thick strands of hair accent this warrior's face.
    (T19 G9: 16).

Figure 3 (a)

VDisc Frame No:	**48605**
Title of Image:	Kneeling Archer
Data of Slide:	1979
Date of Discovery:	1973?
Date of Object:	c. 200 BC
Museum:	Shaanxi Provincial Museum
Pit of Discovery:	Pit #2, Lintong
Color of slide:	Color
Keywords:	Archer,Archer-kneeling,Archer-kneeling-full_size, Archer-right,Archer-side,Archer, terracotta,Warrior,Warrior-kneeling,Warrior-right,Warrior-side,Warrior-terracotta,Hand,Hand-right,Hand-left,Hand-terracotta, Feet,Feet-terracotta,Armor, Armor-terracotta,Pit #2, Qin Terracotta,Qin,
Reference Source:	Capon, Edmund. *Qin Shihuang Terracotta Warriors & Horses*. Australia: Wilke and Co. Ltd., 1982. p. 58.
Description:	All the kneeling archers were recovered from Pit #2 and were very similar, however, they each contained slightly different detailed variations. This figure adopts the characteristic pose of resting the right knee on the ground and keeping the left knee raised. The figure's right arm is held to the side of the right thigh with its hand open for holding a weapon. The left elbow rests on the raised left knee with the hand extended across the chest in order to hold the crossbow. Particularly characteristic of this figure type is the straight, almost arched back. This pose emphasizes the impressions of concentration and discipline. The eyes look directly ahead, while the head is held firmly. Wearing plated armour on the upper half of the body together with shoulder pieces, the archer is only partially protected. The battle robe beneath the armor is distinguished by a series of pleats and folds as it rests over his legs.

Figure 3 (b)

Figure 3. Original image cataloging and indexing records of *The First Emperor of China*
Figure 3 (a) is the original format, while that of 3 (b) has been reformatted

4.2. The CMNet Prototype's Creation Process

• **Preliminary metadata considerations.** Realizing that metadata will be one of the major research areas for *CMNet*, particularly for those collaborators and affiliates from the library-related fields, early discussions on this subject were made with a few key people in the US. The raw data of the 1900+ images (in the format of Figure 3a) were initially analyzed, and a web- and VRA-based entry form was created by Zeng. It has rather extensive number of

entries.[2] A subject category list and subject list were also generated.   Clearly, because of The First Emperor of China's resources are mainly multimedia-based, it was deemed to be more suitable to use VRA than Dublin Core at this time. However, this initial web-based form was never used. The reasons are man. These include considerable effort to mount the form on the network, the need to have experienced assistants who are familiar with both the subject matters as well as the use of the system for data entry, and the slowness to complete the data entry over the Web. At this time, we do not have the kind of experienced staff needed. The network responses are slow and the data entry via the Web will take considerable amount of time to complete even one record. The expenses involved will be prohibitive for the project.

- **Mounting Access relational databases to the selected images --** In order to mount the *The First Emperor of China's Insight*™ *Exhibition* system, in addition to select and process the images in acceptable format -- in our case, in compressed .TIFF -- we also need to provide matching descriptive database information for all images. For this reason, Access relational database forms were created to capitalize the existing raw data as well as using the Meta Elements, Subject Categories, and Subject List generated earlier. Then, the original raw data of the 1900+ images were uploaded to the databases.[3] Ironically, these forms also were not fully utilized, and eventually, the flatten database file was found to be most effective in getting the rush work done.   Clearly, this is not what we recommend, but simply to say that one should find ways to utilize the old existing data to minimize the tedious data input task.

- **Image digitization and digital imaging task --** As stated earlier, *The First Emperor of China* has about 5,000 slides included in the Slide Library Chapter of the Videodisc copy. In 1993, when Kodak's Photo-CD technology was introduced, four professional quality Photo-CDs were created for the selected 400 some slides, which are later included in the Multimedia CD product.   For *CMNet,* we extracted these 400 images at the Photo-CD's highest-level of resolution, each is at the size of about 20MBs.   For the rest of the 4000+ slides, we digitize them one by one at 1200 dpi, with each image size over 20MBs, with the largest one about 75MB. For each of the digitized image, 5 derivatives of different sizes -- 1536, 768, 384, 192, and 96 pixels at the long side of the image -- were created after all the necessary digital imaging work, including cropping, rotating, adjusting levels, colors, contracts, etc…

   Because the Insight® Exhibition Package limits the number of images to 250, thus careful selections were made to include the best images that covers all the necessary categories. In addition, in order to demonstrate what is needed for future research in multimedia context-based retrieval, manual value-added efforts were made to combine images of the same theme on one large image (to be discussed further in the next section).

- **Digital videos --** In addition to database information, text description notes, additional hyperlinked information resources in various formats -- PDF files, URLs, notes, etc. -- 10 segments of digital videos in QuickTime formats are also taken from *The First Emperor of China*'s multimedia CD for mounting on the Insight™ exhibition system.

4.3. What Can the *THE FIRST EMPEROR OF CHINA* EXHIBITION Offer?

In a short one-month time, just prior to NIT 2001, I managed to complete all imaging work for the selected images for *The First Emperor of China Exhibition* by integrating these digital

---

[2] I am grateful to Dr. Marcia Zeng of Kent State University for her preliminary efforts in the metadata area and her preparation of VRA-based data entry forms. These have made it possible for me to make more intelligent decisions for the creation of needed image data for the prototype system.

[3] I am also grateful to Dr. Jian Qin of Syracuse University for her effort in creating the Access relational database input forms based on the existing raw data and the forms from Dr. Zeng, and to Dr. Xia Lin of Drexel University for uploading all the 1900+ original raw data for use in the Access databases.

images with associated descriptive information for a networked environment. *The Exhibition* will enable distributable and scalable online access to images from anywhere. It will provide a dynamic and interactive user environment where one can view, compare and organize images, or even present them to others. The functionalities are summarized in Figure 4:

file				
edit				
group	search		double	select
data	by title		clik to	pan
search	**by category**	**terracotta warriors**	**view**	**zoom**
print	by keyword		image	maximize
return	by period			data
help	by data fields			close
exit				print
				help

Figure 4. Functionalities of Insight™ system

At the writing of this paper, the system has not been completely mounted. Thus Figure 5 is a constructed version of what the Group Window of *The First Emperor of China Exhibition* will look like. All functions for finding, displaying and using images begin from this window. When a given image is selected -- in this case, the "16 back views of the warriors' hair styles" -- the image's descriptive data is retrieved, shown on the upper-left portion of Figure 6 and the chosen image in bigger than the icon in shown in the lower-left of the Figure.
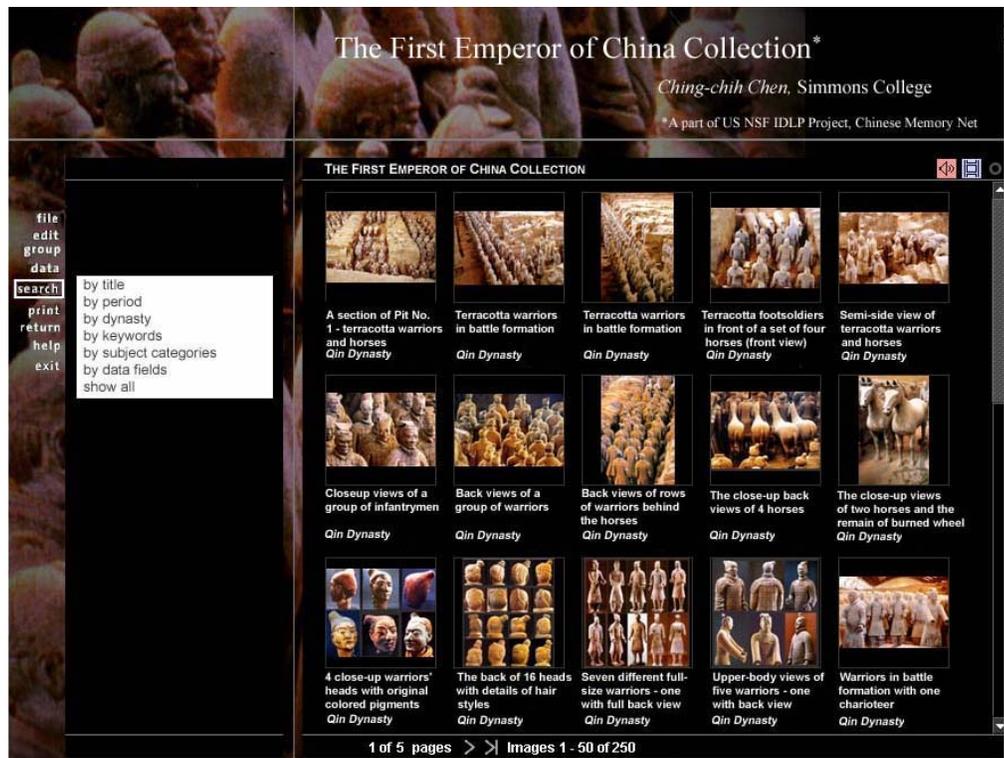
Figure 5. The Group Window showing the selected images of The First Emperor Collection *(Color)*

Double click on the image icon, one will move to the Image Workspace (lower right of Figure 6), where one can view the active image in multiple ways as shown in the function panel in the lower right of the Figure. This value-added image is manually created and is more than 55 MB. Instead of 16 separate images, it is grouped in ONE, thus one is able to gain lot more information then simply zooming or panning of one simple high-resolution image. Thus, Figure 6 shows that when one makes the Selection 1, the enlarged image reveals interesting

"seal-like" image, thus, the Section 2 can further select the area and zoom in to find more details. The color plates in the front of this book can reveal more dramatic information.

The data information shown on the upper left of Figure 6 provides needed descriptive information of the chosen image with as much relevant information as one has. The upper right icons indicate that digital videos related to the image as well as more textual information can be provided with hyperlinks, and notations by users can be made.
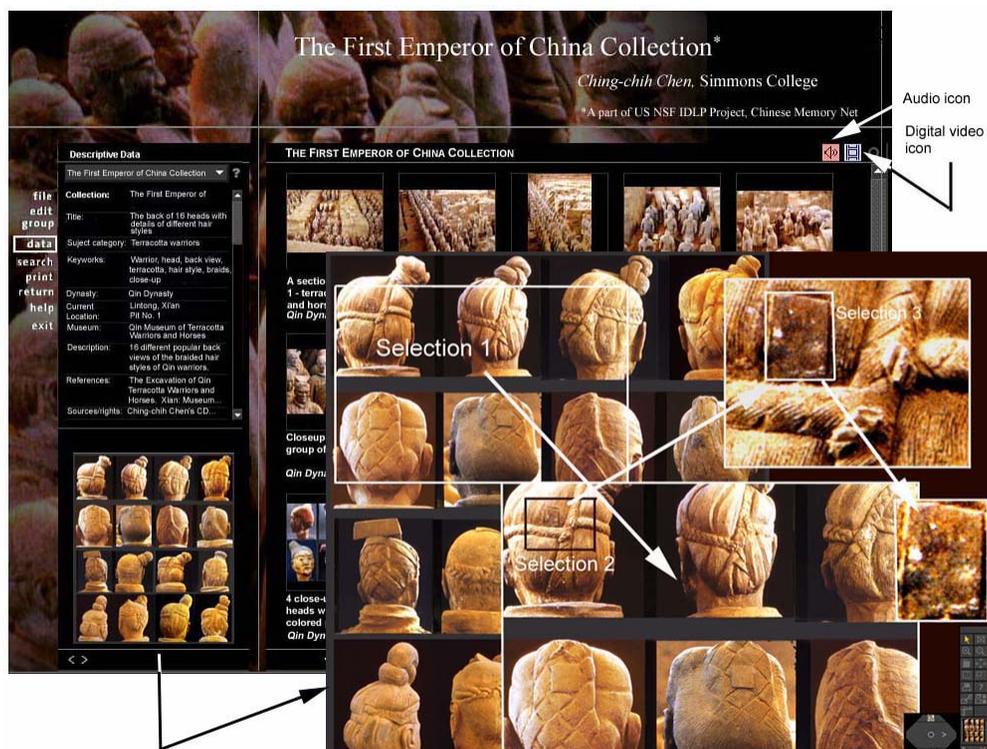

Figure 6. The Image Workspace showing how the selected image of The First Emperor Collection can be zoomed and studies *(Color) -* See color graphic for better illustration

6. LESSONS LEARNED THUS FAR AND RESEARCH AREAS REVEALED

It is an incredible experience to create a prototype of this complexity in one month. Clearly the process was intense and "painful". Yet, we have learned a great deal. During the process, we learned that, first of all, one should value ones old data and try to reuse them fully. Do not start all over again. This prototype would not be possible if we had to create the database information for the selected images from scratch. With the inviting features of *The First Emperor of China Exhibition*, we know more now what we wish to have.   These are what we will need to have further research. There are many wishes and questions, some of which are listed in the following:

- Metadata semantic cross-walk capabilities are attractive, but how good are they? How to design a system that can incorporate and integrate descriptive databases of any structure and from any sources easily?
- *The First Emperor of China Collection* is hopefully only the first of the many autonomous, and heterogeneous systems of *CMNet*, each of which has its own invaluable contents. How to develop a suitable architecture for these distributed as well as local systems will be a vital question.
- With the Emperor's rich database resources, how to utilize that for more productive ontological research which will enhance more intelligent retrieval, translation, etc.

- How to design and implement a multilingual system by decreasing the language barriers with fast retrieval of both English and Chinese data and information resources (Chen, p. 33)
- How to implement true content-based image retrieval (Gao, et al., p. 71) Emperor prototype provides incredibly large number of images of the same shape, form, and color, all with specific meanings. How can they be utilized to further this research in order to push forward from the keyword-based retrieval?
- How to incorporate digital videos in a more seamless way? (Wactlar, p. 323)
- *The First Emperor of China Collection* will have substantial historical geographical information from the Great Wall to the warring states and to the silk road. How to integrate those with the existing system? (Smith, p. 287)
- How to resolve the interoperable problems with this Emperor system with others distributed systems.
- What will be the scalability problems when the number of high-resolution images and digital videos grow from thousands to hundreds of thousands when true internationalization becomes a reality with multi-country distributed systems integrated together with contents and technologies. Is Open Archives Initiative a good start? (Fox, p. 59).

6. CONCLUSIONS

The vision for universally accessible collections of human knowledge as articulated in PITAC's Panel on Digital Libraries Report to the President (see Appendix 1, pp. 559-586) is:

> *"All citizens anywhere anytime can use any Internet-connected digital device to search all of human knowledge. Via the Internet, they can access knowledge in digital collections created by traditional libraries, museums, archives, universities, government agencies, specialized organizations, and even individuals around the world…"*

How can this vision be realized if we cannot even creating a functional and operational global collaborative distributed system on such a minute knowledge base of human knowledge like *The First Emperor of China*. This is why *CMNet* is only contributing to work TOWARD a GDL in Chinese Studies first. Hopefully the experience and knowledge gained will be transferable to related to other segments of human knowledge. *CMNet* is truly a NET, because it is working in a networked environment, and creating a synergetic informal network of researchers on common research interest.

While techniques, tools, and technologies are essential, we should never forget that CONTENT is still the key to any kind of libraries -- traditional, analog or digital, because that is the essence of "human knowledge." So, in developing our digital libraries, let us try not to do less by doing more with our preoccupation with the technologies (by making a big deal of the technology). Let us remember that technologies are tools. Let us also concentrate more on the global content building by pulling distributed resources together using the current technologies, so that there will be something worthwhile to be retrieved when the technologies and methods are in place. This is why it is worthy for us to pursue the "global digital library" development that I have advocated for more than a decade. This is also why we need to start the "million-volume" Universal Library as advocated by Raj Reddy, the PITAC co-chair (Chen, 1999).

We are at the dawn of a new millennium, facing with the ever-expanding "digital" opportunities and challenges. Indeed, this is a magic moment, but like all moments it will not last forever. We must make the most of it. Let's work collaboratively to make this time of change a moment of dazzling opportunity for the global digital library development.

REFERENCES

Chen, Ching-chih. (1994). "Multimedia and the First Emperor of China: Moving toward a digital knowledge base," *Multimedia Today*, 2 (2): 68-71.

Chen, Ching-chih. (1999). "NII, GII, I2 and IT*2 Initiatives: Implications to the digital library development in the US," in *IT and Global Digital Library Development*, edited by Ching-chih Chen. Newton, MA: MicroUse Information. pp. 49-64.