

***NSF Post-DL Futures Workshop***  
Chatham, MA, June 15-17, 2003

***Toward a Global Digital Library***

***from***

***Ching-chih Chen***

Graduate School of Library & Information Science  
Simmons College, Boston, MA 02115  
Chen@simmons.edu

In February 2001, the PITAC's Digital Library Panel submitted its report, entitled *Digital Libraries: Universal Access to Human Knowledge*, to President Bush, in which the Panel stated the vision:

*“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. These new libraries offer digital versions of traditional library, museum, and archive holdings, including text, documents, video, sound, and images. But they provide powerful new technological capabilities that enable users to refine their inquiries, analyze the results, and change the form of the information to interact with it...*

*Very-high-speed networks enable groups of digital library users to work collaboratively, communicate with each other about their findings, and use simulation environments, remote scientific instruments, and streaming audio and video. No matter where the digital information resides physically, sophisticated search software and find it and present it to the user. In this vision, no classroom, group, or person is every isolated from the world's greatest knowledge resources.”*  
[PITAC, 2001].

Clearly we are a long way to go from realizing this vision. Yet, after a decade of sizable investment in digital libraries by the National Science Foundation (NSF) and other federal organizations with several major government sponsored initiatives, such as NSF's DLI-1, DLI2, and IDLP, indeed we have made considerable progress in addressing a considerable number of significant R&D problems for digital libraries. These include interoperability (metadata and OAI), scalability, and information retrieval techniques of both textual and multimedia resources (images and digital videos) etc... In addition, these US initiatives have created fruitful research environment for global R&D activities in digital libraries as well. After the first ten years, we are indeed at the promising “converging and crossing thresholds,” as stated at the January 2003 NSF report, *Revolutionizing Science and Engineering through Cyberinfrastructure* [Atkins et al, 2003]

This *Cyberinfrastructure* Report is pointing us more toward the realization of the vision stated in the PITAC/DL Panel Report. It envisions a program “to build more ubiquitous, comprehensive digital environments that become interactive and functionally complete for research communities in terms of people, data, information, tools, and instruments and that operate at unprecedented levels of computational, storage, and data transfer capacity.” Instead of the concentration mostly on the computer science related R&D projects of the earlier DL initiatives (specifically DLI-1 and DLI-2 which has come to an end), the *Cyberinfrastructure* Report advocates more focus on the creation of really operational and functional digital libraries in the networked environment. At this “historical” workshop aiming for the development of directions and recommendations for long-term NSF DL research, I am honored to be included, and would like to offer the following thoughts for possible consideration.

### **Specific Essential DL Areas for the Next Decade**

As a library and information scientist, topics of significance to me have always been focused in:

- The effective acquisition, organization, retrieval, dissemination, and use of information resources,
- The transformation from data to information and then to knowledge, and
- **Universal access** to needed information, regardless of formats.

These three bullet dots seem to cover most of the things we do and hope to accomplish in either traditional or digital libraries. For digital libraries, because of the “digital” nature, information resources can be shared over the powerful network. With the innovative use of information technology and the integration of many tools and techniques developed thus far and in the foreseeable future, information provision can be more complete, faster, and broad-based. They can be accessed anywhere anytime by anyone who needs them. Thus, the potential should be great.

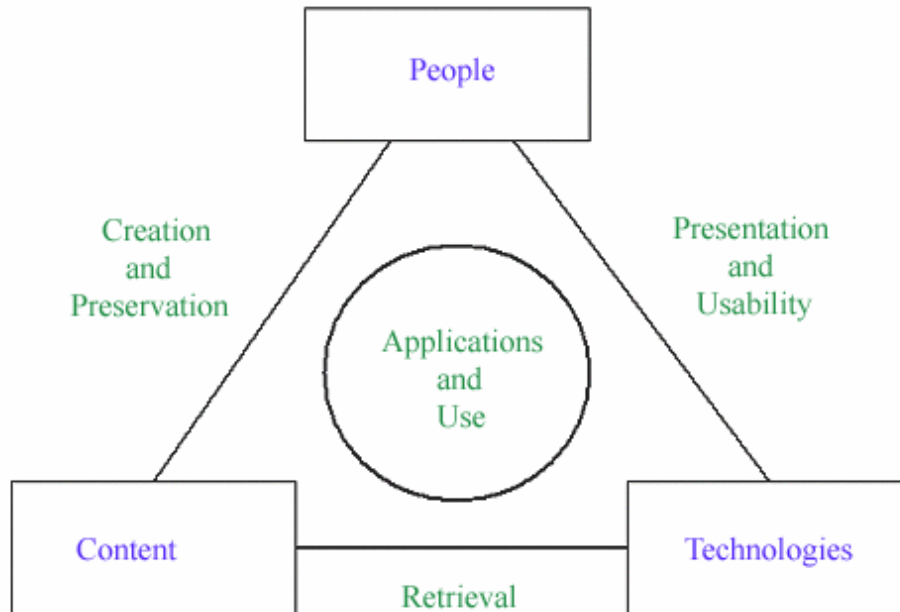
#### *The need for “functional” and “operational” digital libraries*

Yet, from “functional” and “operational” points of views, it seems clear to us that currently, despite of the sizable investment in the last decade, the “digital library” is still a poor shadow of its counterpart, a traditional library. Why is that? This is mainly because most of the DL activities are still sitting at the narrow and specific “R&D” level, and the developed tools have not yet been either utilized or integrated. Also, at the “R&D” level, most research activities are conducted with rather limited amount of useful contents as well as limited metadata and descriptive annotations of these contents.

- The need to integrate technology, content, and users

In order to have more “functional” and “operational” digital libraries, we need to do much more by integrating technology, content and users. The *Report of the DELOS-*

NSF Working Group on Digital Imagery for Significant Cultural and Historical Materials provided a conceptual framework for digital libraries as follows:



This conceptual model attempts to illustrate the relationships among people, content, and technologies in developing research agenda. The Report states:

*“Our interdisciplinary research will develop **technologies** to enhance the way **people** create and access the **content**... People encompass all users, from curators and library and information scientists, to scholars, teachers, and students in all areas of the humanities, to citizens of all cultures. Content is the vast array of significant ... materials throughout the world. Technologies are the enabling research and development in all related technical areas such as information retrieval, image processing, artificial intelligence, and data mining.*

*We recommend focused, interdisciplinary, research programs along the three edges and the center of the triangle, areas that traditional research programs currently neglect. The research area between people and content is the area of digital imagery **creation and preservation**. The area between content and technologies is the efficient and effective **retrieval** of the content using technologies. Research into **presentation and usability** will enhance the ability to access the content. Effective **applications and use** of the research results, under lifecycle management, will integrate research of the three related areas.”*

- *The need for comprehensive metadata*

Although creating metadata and description annotations is a very tedious and labor-intensive activity, yet, it is vitally important and cannot be ignored. In this regard, in

addition to continuing research in areas of interoperability, increasing emphasis on the semantics of and correlative relationships among data should be explored through computer learning techniques, etc. so that effort in creating comprehensive metadata and annotated descriptions can be minimized.

*The need for “quality” contents – Need to address copyright and IP issues*

It is widely known that despite of all the efforts, currently there is a serious lack of quality digital contents. Thus, one of the most significant areas for the next decade will have to be related to quality content development. Quality is to be defined by information seekers as most up-to-date, comprehensive, authoritative, etc. as they see fit. For example, for some countries, although their rich cultural and heritage resources are heavily sought after by others, these resources may not be considered by their government and/or university educators as the most important when compared with the current scientific and technical information resources. Clearly this is tied with the national policies. Yet, these current resources are simply difficult to obtain and be sharable for public free access. One of the most serious barriers is related to copyright and intellectual property (IP). With these issues clouding over the sky, it is difficult to develop large-scale quality multimedia contents for operational and functional DLs. In addressing these “legal” issues, we need to go far beyond the “technological” solutions like digital water mark etc.

*The need for global and multidisciplinary collaboration*

One of the potential remedies for addressing the need stated above is through global and multidisciplinary collaboration by networking various distributed digital contents. This need was advocated with and fully shared by many international DL partners [Chen, 2001]. From R&D angle, this will promote more problem-oriented and synergistically complimentary research in addressing issues such as interoperability, scalability and multilingual; while from the content angle, this will encourage more resource sharing of valuable contents currently housed in different institutions, organizations and countries in the world provided the collaborative partners are willing to share their valuable resources. Currently this willingness is still doubtful, and efforts will have to be made to find ways to create more effective infrastructure and to provide more attractive alternatives for sharing. If this is possible, then large-scale “digitization” will be needed to create much more digital multimedia contents quickly. Projects like the *US-China Million Book DL* and the *US-India Million Book DL* projects, supported by NSF since 2000, have made progress in creating large quantity of digital contents which can have great R&D potential. Yet, “quality” issues remain to be difficult ones. For example, most materials scanned are older materials free of copyright concerns, thus they are mostly historical or materials of native languages. There is a real need for more English-language and current contents which are not readily available due to copyright and IP issues.

In conclusion, it is worth noting that at the latest *DELOS-NSF Workshop on Multimedia in Digital Libraries*, held in Chania, Crete on June 2-3, 2003, similar concerns expressed above were articulated. More international and interdisciplinary collaboration was deemed a must. We have a great future toward a global digital library!

## References

Atkins, Daniel E., et. al. *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue Ribbon Advisory Panel on Cyberinfrastructure*. January 2003. Available online at [http://www.communitytechnology.org/nsf\\_ci\\_report/](http://www.communitytechnology.org/nsf_ci_report/).

Chen, Ching-chih, ed. *Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration*. Beijing, China: Tsinghua University Press, 2001. 614 pages.

*Report of the DELOS-NSF Working Group on Digital Imagery for Significant Cultural and Historical Materials*. [edited by Ching-chih Chen and Kevin Kiernan]. December 2002. [http://dli2.nsf.gov/internationalprojects/working\\_group\\_reports/digital\\_imagery.html](http://dli2.nsf.gov/internationalprojects/working_group_reports/digital_imagery.html)

PITAC. Panel on Digital Libraries. *Report to the President: Digital Libraries: Universal Access to Human Knowledge*. February 2001.