

International Conference on Digital Libraries 2004 Conference

Debal C. Kar and Michael Seadle

Introduction

The International Conference on Digital Libraries (ICDL) took place at the India Habitat Centre in New Delhi, India, on February 24-27, 2004. The Energy and Resources Institute (TERI, <http://teriin.org/events/icdl>) organized the conference in partnership with the Government of India, Department of Culture, Ministry of Tourism and Culture and the Ministry of Communications and Information Technology, the Commonwealth Educational Media Center for Asia (CEMCA) and the Centre for Development of Advanced Computing (CDAC). This is one of a growing series of digital library conferences that bring together computer scientists and librarians. Some earlier examples include the 2000 Kyoto International Conference on Digital Libraries (which Kyoto University, the British Library, and the US National Science Foundation organized), the 2001 Bangalore International Conference of Asian Digital Libraries (ICADL), or the 2002 Singapore ICADL (Plate 1).

With more than 740 registered participants it was clear that the conference had a great impact in the digital library circles the world over. The registered participants also included about 100 from 33 countries in Asian, African, European, American and Australian continents. The 55 invited speakers representing 16 countries consisted of eminent speakers and information scientists in the digital library field. It was also remarked that ICDL 2004 was the biggest ever digital library conference organized so far in the world. Due to constraints of space, registration was frozen a fortnight before the conference.

Background

Digital libraries and knowledge management are innovations, the implementation of which is nascent in the developing countries. However, they hold the promise of becoming key technologies for knowledge creation and management in the future. It is under these circumstances that TERI proposed the organization of an international conference that would deal with all subjects related to digitization and knowledge management.

In February 2003 the Government of India launched the National Mission for Manuscripts, aimed at preserving the valuable manuscripts in India. The mission is designed to undertake elaborate surveys, run awareness programmes, and impart training in conservation and preservation techniques. Since the objectives of the proposed conference and the mission were on the same wavelength, we approached the Department of Culture, Ministry of Tourism and Culture, Government of India, with the proposal for organizing the ICDL 2004. And we are grateful that they wholeheartedly supported the idea.

Objectives

The objectives for organizing this conference were to: bridge the knowledge gaps between developing and developed countries; initiate capacity building activities in digital libraries; evolve a road map for the digitization of archives, manuscripts and libraries; provide a forum for facilitating interaction among participants; and formulate recommendations on digitization technologies, acts, and policies.

Themes

The themes were carefully selected for the conference and included:

- Digital library concepts.
- Digital divide.
- Country initiatives.
- Digital library policies.
- Digital library services.
- Digital library communities and digital library consortium.
- Digital library security.
- Copyright issues and digital rights management.
- Archiving and sharing cultural heritage and history.
- Manuscript: archiving and digitization.
- E-publishing and multimedia.
- Metadata issues.
- Multimedia and multilingual.
- Content management and knowledge management.
- E-learning.
- Semantics, thesauri and ontologies.
- System scalability and interoperability.
- Technology: planning, development, and management.
- Digital library architecture, network and technology.
- Storage and retrieval systems.
- Human-machine interface and User interaction.
- Futuristics.

The following topics and speakers were invited to give tutorials and this captures the range of interests covered at ICDL:

- "Finding information through the Internet and the WWW" by Paul Nieuwenhuysen, Vrije Universiteit, Brussels, Belgium.

Plate 1

Inauguration of ICDL 2004. From left: Professor Hsinchun Chen, Professor, Department of MIS, University of Arizona, USA; Dr R.K. Pachauri, Director-General, TERI; Professor N. Balakrishnan, Professor and Chair, Division of Information Sciences, Indian Institute of Science, Bangalore; Dr A.P.J. Abdul-Kalam, Honourable President of India; Mr Jagmohan, Honourable Minister of Tourism and Culture, Government of India and Mr Vinod Bhargava, Additional Director, Division of Information Technology and Services, TERI



- “Thesauri and ontologies” by Nicola Guarino, National Research Council, Padova, Italy.
- “Institutional repositories” by T.B. Rajashekar, Indian Institute of Science, Bangalore.
- “Building digital library using open source software (GSDL)” by Ian H Witten, University of Waikato, New Zealand.
- “Knowledge management systems: development and applications” by Hsinchun Chen, The University of Arizona, USA.
- “Automated text classification” by Fabrizio Sebastiani, Consiglio Nazionale delle Ricerche, Italy.
- “Dublin Core” by Stuart Weibel, OCLC Office of Research, USA.
- “Web technologies and persistent information and knowledge repositories (XML, RDF & Co)” by Erich J. Neuhold, GMD-IPSI Darmstadt, Germany.

The full conference included 80 papers which were accepted for presentation at the conference. In addition, 75 submissions were selected to be presented as poster sessions. The conference proceedings consisted of three volumes, one for tutorials and the other two for invited and contributed papers and poster presentations. A total of 115 papers and 73 abstracts were included in the proceedings. The proceedings were distributed to the

delegates at the time of registration. (A few copies of the proceedings are available for sale. Please contact outreach@teri.res.in or icdl2004@teri.res.in for details.) There were product presentations and two dozen exhibitors.

During the conference daily bulletins were distributed noting discussions in each topical track.

Michael Seadle writes of his experience at ICDL:

I attended this conference as an invited speaker, along with a number of colleagues from Germany (Peter Schirmbacher from Humboldt University, Dieter Fellner from Braunschweig, and Eric Neuhold from Darmstadt), Austria (Andreas Rauber from the Technical University of Vienna), Switzerland (Jean-Marc Comment from the Swiss Federal Archives), Australia (Simon Huggard from Monash and Liddy Nevile from La Trobe), New Zealand (Ian Witten from Waikato), Britain (Norman Wiseman from Nottingham), and the USA (Stuart Weibel from OCLC, and Ching-chih Chen from Simmons). This is far from an exhaustive list, but includes most of the people I interacted with on a regular basis. Many of us stayed at the same hotel at the Habitat Centre and spent time together at lunches and dinners. Conversations with them ranged from discussions of ProPrint (<http://edoc.huberlin.de/proprint/>) to Greenstone (www.greenstone.org/cgi-bin/library)

to possible speakers for other conferences. The real value of these conferences is often the face-to-face interactions with knowledgeable colleagues, and ICDL offered rich opportunities.

At one point we had an entirely German-speaking table (myself included) where everyone could talk freely without danger of inadvertently giving offence to our hosts. One colleague said there were really two conferences taking place: one that was a training session for local attendees, and another for invited speakers that was more like the US Joint Conference on Digital Libraries or the European Conference on Digital Libraries. To some degree this is probably always true at large technology-oriented conferences, where the speakers will naturally be experts who probably know each other, and the remaining attendees are people who have come to learn. This was clearly not an Indian-European divide, since the speaker list included impressive Indian colleagues like N. Balakrishnan (Bangalore and Carnegie Mellon), Harsha Parekh (SNDT Women’s University, Mumbai), and Om Vikas (head, Computer Development Division, Ministry of Communications). Nonetheless fairly basic questions tended to dominate the discussion periods. While the same is true at many US-based meetings where a large number of people come to

conferences to learn, cultural differences made it more obvious in Asia.

Another colleague with whom I shared this dual-conference remark said that he particularly appreciated the discussion period because the interactions were lively and the people at the conference really wanted to learn as much as possible. He said the dialog was more stimulating than with many of his students at home, who sat in silence and were too embarrassed to ask about what they obviously did not understand. He came expecting something more like Computers in Libraries or similar gatherings with a big training component and less peer-group discussion, and he found it delightful.

The inaugural address was by His Excellency Dr A.P.J. Abdul Kalam, President of India (fulltext of the President's speech is available at <http://presidentofindia.nic.in/scripts/sllatest1.jsp?id=282>). The conference ran with admirable efficiency despite some unexpected twists. When the President of India, Dr A.P.J. Abdul Kalam, addressed the conference as planned, he added a request for recommendations about a wide range of digital library policy matters. And he wanted the recommendations before the conference ended. The organizers hastily put together lists of experts in particular areas, and made mid-session announcements asking certain people to identify themselves. My name was on the list, and I found myself unexpectedly led off to a conference room where a discussion was already underway. I joined in, of course, and eventually asked why I was there. The answer was that the organizers had put me on the committee to make recommendations about copyright and intellectual property. In fact, I ended up as convener of the group. Other members of the committee included Harsha Parekh and Om Vikas.

Working out reasonable copyright recommendations that would facilitate digital library development in India without undermining India's membership in the Berne convention took much of my remaining time, and eliminated any temptation to slip in some local sightseeing. In compensation it provided a unique opportunity to work with remarkably

able Indian colleagues on issues of immediate significance. They (and the President) were clear about India's need for access to the same intellectual property as scholars in Europe and the USA, but at prices that reflected the purchasing power differentials. They needed to consider both mechanisms to give reasonable access to intellectual property produced elsewhere and to open Indian-produced intellectual property internally to appropriate educational uses. Some of the committee recommendations included:

- *Registry.* An online registry should provide every digital work produced in India with a digital object identifier.
- *Depository.* A legal depository requirement would help to determine authenticity and ownership as well as preserve India's intellectual heritage.
- *Copy-left.* A statutory copy-left provision could enable a rights owner to surrender exclusive rights after a certain period.
- *Compulsory licensing.* This has already been used for cable television in India, and could be extended to digital applications.
- *Moral rights and ethics.* These can help to ensure that digital works are handled fairly – Teri, 2004 (my paraphrasing).

It is hard to predict what will come from these recommendations, but even if they merely increase the the Indian Government's consciousness of digital library and intellectual property issues, ICDL will have had a significant result.

Some excerpts from the President's address which provide insights into his priorities include:

In the twenty-first century, a new society is emerging where knowledge is the primary production resource instead of capital and labour. Efficient utilisation of existing knowledge can create comprehensive wealth for the nation in the form of better health, education, infrastructure etc. for improving the quality of life. Ability to create and maintain the knowledge infrastructure, develop knowledge workers and enhance their productivity through creation, growth and use of new knowledge will be the key factors in deciding the prosperity of this Knowledge Society.

Digital library is an important component for capturing the explicit knowledge. This has to be supplemented with the implicit knowledge to the digital library system, which will eventually get transformed into a knowledge management system.

The future Digital Libraries would have speech interface. Though this technology world over is progressing, still many more things have to be done. I would urge the learned audience to work in close collaboration to ensure introduction of Natural Interfaces to the Digital Libraries world over. It is important that we take on this mission of integrating all forms of knowledge and culture into our Digital Library.

The digitization initiative has picked up momentum in the country; this is the time we have to make consistent national policies and procedures, which will lead to effective management and control of the data leading to enhancement of national knowledge base like what we have done in IT. Policy makers should take into account of the standardization requirements, inter-operability, copyright issues, classification of documents and selection and use of number of library information systems available with various organizations in the country in different standards. I would recommend deliberation of this issue in the conference and constitution of a multi-disciplinary task force for working out the draft policy document for implementation. The policy should take into account the dynamics of the Web services technology and keep provision of on-line improvement and variation in future digital library systems. The task force should have access to all available resources so that they can ensure prevention of duplication of efforts.

We have to create a Knowledge Management Grid with the Central Digital Library Data Center equipped with the comprehensive Virtual Digital Library and Knowledge Management System into which all the participating organizations are connected with broadband along with Internet connectivity.

For enabling knowledge connectivity in our rural areas, we need to have a comprehensive plan for developing new infrastructure for extending the digital library services in regional languages. These include – development of Optical Character Recognition (OCR) software in all the Indian languages, language independent operating system, database

servers, search engines, Web servers and messaging servers. This will enable the digital library initiative to percolate to the rural masses in the form of e-governance, tele-education and tele-medicine. This has to be done on a mission mode with the active participation of Government, educational institutions, R&D organizations along with private sector enterprises.

It will be useful to digitize and store textbooks in the library, which can be accessed by the students whenever required. This additional facility will enable easy availability of books to the students for instant reference and study through Internet and intranet access. We need to integrate the tele-education system with the digital library, so that the students can read and refer to the books suggested by the teacher from a distant location through online e-learning services.

There are large numbers of files and records occupying valuable space in our State and Central Government offices. It is essential to segregate the important files which have to be preserved for a long duration. These segregated files have to be digitized and stored in the digital library.

Land records have to be digitized and verified with satellite imagery and stored, which needs to be linked with e-governance applications for issue, transfer, conversions and additions and deletions. It should be linked with the revenue collection, estate management, and municipal records. Digitization and information flow has to go parallel to get the real advantage of e-governance workflow in the land records and its management domain.

Digital data of a voter should be made available on-line over the Internet for all the States and Union Territories as a digital book for reference. Each State should have a dedicated site for presentation of the data accessible to all voters as well as election officers.

In this millennium when the rate of flow of new books/journals has increased substantially, there is a need to have a re-look at the lock-in period of copyright documents. This International Conference could initiate action for reducing the copyright duration substantially.

We should see that all the schools, colleges, and universities digitize their libraries in their own native languages

and connect to the outside world within the next four years. We have to ensure availability of fiber optic cables, satellite communication and wireless infrastructure especially in remote areas. It is also essential to realize high bandwidth technology like Multiple 10 Giga Bits connectivity across the country.

Keynote addresses

There were three keynote addresses delivered and accompanying notes highlight the content of each:

(1) *Mr Kiran Karnik, President, NASSCOM.* Digital library in a broader context is nothing but a database. The objective of the digitization process should be "empowerment of people". The Indian information technology (IT) industry has done remarkably well globally in recent years. But, its orientation being towards export, it has given the industry a negative image within the country.

There are technologies available, but use of the appropriate ones, not necessarily high-end ones, is the key. While conceptualizing any digitization project, it is important to have some measurement criteria for outputs as well as outcomes. Outputs should be in terms of number of documents digitized; outcomes should be in terms of ease of accessibility, the spread or the reach, and the ability to access information and knowledge. Both qualitative and quantitative goals should be incorporated into the project.

Another important dimension of the digitization process is the coordination between information technology and library science professionals. These professionals will have to work as a team to achieve the desired objectives.

A few successful examples from India were cited. In Madhya Pradesh, in an attempt to empower the rural populace, several kiosks were set up to provide information. These kiosks are totally entrepreneur-driven and a small fee is charged for their use. In Karnataka, about 20 million land records were computerized, bringing in more

transparency and authenticity of ownership. In Andhra Pradesh, through the e-Seva project in Hyderabad, electricity bills, telephone bills, property taxes, etc., can now be paid through the Internet. In Rajasthan, many government policy documents and others have been digitized and they are accessible to the public.

(2) *Professor Ching-chih Chen, Simmons College, USA.* Whereas digitization and digital libraries have closer a connotation to storage and retrieval of text and images, digitization has immense potential to facilitate access to and study of cultural artefacts, be they stored in museums or in actual use in communities spread the world over. The Project Emperor 1 in China was the first major project of this kind and captured 108,000 pictures on a videodisc. Though this was offline, with the spread of the Internet, such contents can be made available to any corner of the world. Universal access implicitly assumes a largely unidirectional flow of information from the developed world to the developing world but the latter is often richer in traditional wisdom, crafts, and cultural artefacts – digital technologies have the potential to preserve that heritage and make it accessible. The concept was neatly illustrated through a triangle with people, content, and technologies as its three vortices. There is a wealth of material (content) and technologies to capture, store, and distribute it; what is needed is the willingness to share. The talk elicited queries from members of the audience about image-based retrieval versus metadata-based retrieval, about appropriate changes in the present curriculum of library and information science education, and about a national policy on sharing content.

(3) *Professor N. Balakrishnan, Indian Institute of Science, Bangalore.* If technology push, crashing costs, and global connectivity are the signposts for the coming decade, digital libraries are the next step in the convergence revolution. The overall goal is to have all public knowledge online, available for

free to all, everywhere. At present, there are only 50,000,000 books in the world and a few billion dollars could bring these online. However, there are several obstacles to reaching the goal of universal access. Use of copyrighted material, privacy, reliability of information, and, most important, the cost-effectiveness of the entire operation are still issues to be discussed and resolved.

In the twentieth century, governments created British Broadcasting Corporation (BBC), Public Broadcasting Service (PBS), All India Radio (AIR), and also the public library system in the interest of public good. These provided compensation for artists and writers while providing free access to the public. The total global expenditure in public broadcasting and public libraries exceeds a hundred billion dollars. Our ancient kings supported all poets and scholars so that people at large can benefit. It is time to find the twenty-first century equivalent of BBC, AIR, and PBS.

Perhaps the solution lies in the four Cs (Consortium for Compensation of Creative Contents), which suggests setting aside 25 per cent of the current national expenditure on public broadcasting and public libraries. Authors are encouraged to put their work on the Web after a few years of commercial exploitation and in return may be exempted from tax.

Knowledge multiplies whenever bits are circulated on the Web. The Internet and the four Cs are necessary to push each other towards the exponential increase in connectivity and knowledge. Technology follows the law of accelerating returns, not the law of diminishing returns. The "expanding universe theory" means that the time span of a technology will be shortened.

It is impossible to summarize all the papers delivered in the various

tracks. Suffice it to say that interesting, thoughtful and important comments were shared and that the conference proceedings will have to be consulted for specific details.

Conference outcomes

In his inaugural speech, the President expressed the opinion that:

...we have to make consistent national policies and procedures, which will lead to effective management and control of the data leading to enhancement of national knowledge base ... Policy makers should take into account the standardization requirements, inter-operability, copyright issues, classification of documents and selection and use of a number of library information systems available with various organizations in the country in different standards.

He further recommended:

...deliberation of this issue in the conference and constitution of a multi-disciplinary task force for working out the draft policy document for implementation.

Such a group was formed and began meeting during the conference. Four sub-groups were also constituted to deal with resources; technology; users; and policy and management. Another sub-group on economics and sustainability will be formed later and the substance of their charges are as follows:

- *Sub-group on resources.* Identification of resources need to be undertaken under cultural and literary heritage of India; Indian history and science; environment; tourism, folklore and music; adult literacy and government publications; demography; learning; and medicine.
- *Sub-group on technology.* The main task of this sub-group will be to develop suitable technology for adoption for the task at hand.

A technology should be developed for each resource.

- *Sub-group on users.* This group will identify various users and stakeholders and their needs. The users include activists from all walks of life. The group identified about 20 types of users and their requirements.
- *Sub-group on policy and management.* This group will deal with policy related matters, including intellectual property rights, etc.

It was intimated that an amount of 2.50 billion rupees has been set aside for the above activities for the next three years. Work on the aspects mentioned above should start immediately.

The project intends to take at least 1.5 million documents of approximately 200 pages each on to the Web during the period. Adequate storage facility should be provided with moderate connectivity between resource centres and users.

Michael Seadle states that:

I cannot conclude this report without some praise for our hosts. A certain natural chaos exists in India, but did not exist at ICDL. Airport pickup and delivery was handled with perfect punctuality. Internet connections were arranged when requested. Printing problems were solved. Last-minute PowerPoint changes were incorporated. Meals for over 750 delegates appeared and were cleaned away without delay. I have attended conferences in Europe and the USA where those elementary mechanics worked less well. Dr Debal Kar and his team did an impressive job.

Debal C. Kar (dckar@teri.res.in) was the conference coordinator of ICDL and *Michael Seadle* (seadle@msu.edu) is the Assistant Director for Systems and Digital Services, Michigan State University Libraries, East Lansing, Michigan, USA and Editor, *Library Hi Tech News*.