

Conference Bulletin

International Conference on Digital Libraries 2004

Knowledge creation, preservation, access, and management

Organized by TERI in partnership with Department of Culture, Ministry of Tourism and Culture, Government of India

New Delhi • 24-27 February 2004

<http://www.teriin.org/events/icdl>

In brief . . .

KEYNOTE ADDRESS 2

Chair

Prof. V S Ramamurthy, Secretary,
Ministry of Science and Technology,
Government of India

Speaker

Prof. Ching-Chih Chen, Professor,
Graduate School of Library and
Information Science, 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 off-line, 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 tech-

nologies 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.

SESSION V

Session VA

Country Initiatives

The integrated digital science library of the Chinese Academy of Sciences aims to make available at every researcher's desk-top exhaustive databases pointing to information held in all major areas of STM (science, technology, medicine), including those that cover published literature on a common platform. The digital library has also established a reference service to answer users' queries online. The integrated library currently combines the resources of over 70 libraries. Besides being a repository of knowledge, the library adds to it by commissioning its own research on such topics as intellectual property and digital library standards. A cautionary note was introduced in importing STM databases: 'modify your library's policy of collection development and know how to evaluate them.'

Equally salutary were the lessons learnt in another digitization project: (1) outsourcing can prevent the loss of productivity due to staff turnover, (2) it is better to scan all images to a uniformly high resolution, and (3) power spikes can be a major setback. The digital library movement may originate with the objective of serving users in a particular country – be it a small country like Nepal or a large one like Iran – but universal access will be the eventual goal of the movement.

Session VB

Content Management and Knowledge Management - I

The discussion over the definition of 'IM (information management)' versus 'KM (knowledge management)' is never-ending and overlapping, but its success (whether of IM or KM) will largely depend on how efficiently we can organize, and provide access to, all the different varieties of information that are required for the day-to-day business operations as well as for making strategic business decisions. And, there are various approaches to organizing information. E-journals present new integration challenges for libraries and have certainly been a value-added service. E-journals support paper-based and electronic methods of knowledge dissemination, provide remote

access, permit simultaneous use by more than one user, provide instant access, offer search capabilities, and provide such feature as links. Knowledge management is not a new concept to India; the 'gurukul' system, the oral tradition of transmitting texts, and such treasures available today in the form of manuscripts as the *Vedas* and the *Puranas* are evidence enough of that. Digital technology is a powerful tool to disseminate, present, and possibly preserve this heritage. Economies the world over are moving from being industrial economies to information economies and finally to knowledge economies. Transforming information and knowledge into intelligence is the fundamental basis of competition to achieve excellence and world-class performance.

Session VC

Semantics, Thesauri, and Ontologies

The session focused on the role of thesauri and ontologies for searching and accessing desired information. It is now believed that only word-based approach cannot yield precise and most-desired information. Lexical knowledge; synonyms, contextual mapping; and multi-linguality aspects together play a role in ontology-driven information access and knowledge retrieval.

Being a heterogeneous collection of data in multiple forms, a digital library should be annotated. Various graphs, acyclic ordered annotation, and key graphs may be deployed. Ontology maps an object to concepts, which is represented via symbolic language to achieve language-independent retrieval. Various technologies of ontology representation were suggested in the session, and XML standard with variations specific to digital libraries were presented.

SESSION VI

Session VIA

Digital Library Policies

As the session devoted to policy-related matters, the emphasis was naturally on broader issues and not the operational details of digitizing. In focusing all available resources on meeting technological challenges, the significance of cultural and organizational issues is sometimes lost sight of. Just as technology has brought people and resources closer, so it can separate them because of incompatible technical standards. Planning digital libraries at the national level requires a holistic and long-term perspective, participatory approach, and commitment; without that, technology can waste precious resources. 'Assessing and preserving our past for the future depends on the decisions taken today.' At the local level, the head of a library needs to appreciate that the library's mission and goals should be in harmony with those of the organization of which the library is a part. This understanding acknowledges that planning a library is rooted in the policies of the academe. As the Internet and digital documents become part of everyday life, users increasingly tend to believe that any workstation is as good as any library: it is up to librarians to prove them wrong by making users aware of the added value that a library offers through careful selection and organization of knowledge.

Session VIB

Content Management and Knowledge Management - II

The Internet is changing the way we live and do business. It gives opportunities for libraries, governments, and businesses to deliver information and services

more efficiently and to interact with stakeholders: citizens, patrons, other businesses, and other government agencies. Digital governance makes the government a major user of information technologies, a collector and maintainer of very large data sets, and a provider of critical and often unique information services to individuals, states, businesses, and other customers. The need for language-centric research in India is ever increasing because of the needs of a multilingual society. Information overload, greater demand for quality information, and movement towards knowledge-based wealth creation are some of the reasons that give rise to the need for a KM-based digital library. The ACADO digital library software developed by IITM-K is a popular KM-based digital library software. Local commercial banks in Malaysia, by adopting KM, have increased knowledge sharing horizontally (through departments, and functional and business units) and vertically (through tiers of organizational hierarchy). The integration of KM with other existing enterprise systems in an organization, data integration, and other technological complexities make the implementation of KM in an organization a challenging task.

Session VIC

System Scalability and Interoperability

In the session, while discussing system scalability and interoperability, statistical methods to achieve machine-aided translation were explained with the help of sample results. Speakers talked about metadata schemes and interoperability versus localization. A possible model to enhance interoperability and localization of metadata and metadata

schemes was deliberated upon. Basic concepts of the Dublin Core metadata, Warwick framework, layered model for metadata, and registry were explained. It was stressed that no single metadata set is sufficient; a layered model with a combination of various approaches is needed. Besides, mediation, cumulative information model mapping method, OAI-PMH, and digital library network were covered in the discussion.

SESSION VII

Session VIIA

Archiving and Sharing Cultural Heritage and History Globally

Preservation is about linking the past with the future. Technology is fast changing. And digitization as a preservation option is not yet tested. Hence, it is safe to have multiple preservation options of data. A hybrid model of microfilms and digitization with fungibility option is ideal in the given context: microfilms could be converted to a digitized version and vice versa. Another option is to maintain the microfilmed and digitized versions together. While digitization has tremendous advantages, it is a huge task that involves managerial and labour resources and technological involvement of a large number of partners. The Russian example juxtaposed to the Albanian one shows how complex the former one turned out to be, spreading over a decade. Maintaining a standard protocol such as International Standard for Archival Digitization (General) was also difficult as the manuscripts varied in structure and character. The developing nations, when they attempt digitization, must have adequate financial resources lined up. Preservation of images like the Beijing Alleys would be ideal if they can be

used for multiple applications such as sociological research, community study, and tourism industry.

Session VIIB

e-Learning - I

The session started with a presentation on the Tamil Virtual University, which has emerged as an effective online platform for meaningful education. The online university offers various levels of courses for children and is a meticulously-organized repository of knowledge, with audio-visual tutorial tools and a user-friendly search facility.

Technology dictates the pace at which digital libraries are developed but the manner in which technology is used depends on information professionals. A web-enhanced delivery model for information dissemination was discussed, and the development of a scalable, highly interactive model for need-based training in future was stressed upon. The existing network of distance-education institutions – such as the Indira Gandhi National Open University – with their available infrastructure could be used to realize the targets.

The success of the AVU (African Virtual University) was cited to highlight the benefits in integrating ICTs (information and communication technologies) in information services. The AVU's electronic library support service has shown the way for other institutions to enhance the quality and usefulness of their services. Modes of cost-benefit analysis and evaluation of e-learning projects were discussed. Feedback from users was considered to be a reliable index to measure the success of such efforts.

Session VIIC

DL Security

Reliable 'access management' was projected as a key to develop security in digital libraries. From the security point of view, attribute-based approach is better than identity-based approach. New technologies do facilitate new data management techniques and accessibility methods but pose enormous threat to the digital libraries in many ways. There are many weaknesses with regard to these technologies; TCP/IP, configuration, and network policy often result in security problems. Implementation of firewalls and additional securities were put forward as a solution to these threats. It was also suggested to put in place a disaster-recovery plan for all digital libraries to meet out emergencies such as data loss due to unforeseen reasons. When we further move up, the network traffic would also move up, which could again cause security flaw. To overcome this, a mesh topology – even for nodes – was suggested. Implementation of security policies for every organization, and following and modifying that at regular frequencies was strongly recommended during the session.

SESSION VIII

Session VIIIA

Futuristics

Digitization will be a leveller in terms of democratization of information. However, it could also offer unequal access to information. This is because of the varying penetration levels of digitization among societies. One of the studies found that Internet penetration in the USA in 2001 was as high as 50%. Comparable numbers for India and countries

in sub-Saharan Africa were 0.7% and 0.3%, respectively. The future of the books will be in the order 'books to bits to e-books'. The computer science department at the University of Watako, New Zealand, has already begun working on some futuristic projects. These include realistic documents that enable users to browse and flip through pages of books by clicking the mouse, a book without words (for first-aid purposes), and hierarchical phrase-browsing. As digitization becomes the norm, machine translation will play a crucial role in digitizing large volume of data lying in all parts of the world. One of the challenges then would be to locate the needed information. Also it will be a challenge to access such information in one's own language. For these purposes, a general-purpose search engine or machine translation systems that are domain-driven need to be developed. Countries like India need a multi-language open-source software and should also work according to the principle 'rise, raise, and race', realizing the short time left to bridge the digital divide.

Session VIII B *e-Learning - II*

Technology, by itself, is morally neutral and does not transmit

values: it is a vehicle for whatever contents we choose. ICTs offer a great opportunity to remove barriers in education and to facilitate learning. However, there still persists a wide gap between the potential of ICTs and the effectiveness of their use. There is a need to integrate the use of ICTs with formal education systems, since the notion of formal education also works as a barrier to acceptance of e-learning by a learner.

SCRAN (the Scottish Cultural Resources Access Network) – a nationwide consortium of museums, libraries, and archives – was cited as an example of a massive project that sustained itself even two years after the grant ran out. Flexibility in adapting to the technical requirements of its target users, combined with aggressive marketing, has resulted in SCRAN being licensed to every publicly-funded school in Scotland, several in England, and over 90 educational institutions worldwide.

One significant manifestation of electronic publishing is the arrival of e-books. However, there is an air of uncertainty on part of publishers as regards the nature of the products and the market in the coming days. Citing the evolution of the e-Book Working Group – a UK national body assigned to formulate a national strategy to

address the issues – the session provided insights on devising frameworks, which various countries could consider for adopting.

Session VIII C *Human Machine Interface*

Panellists discussed hierarchies generated by humans to humans and machines to machines. The automated text categorization is essential for organizing digital libraries. Programmes capable of assigning one or more categories from a predefined set based on content of the document would lead to construction of automatic text classifiers, and as a result, formation of text categorizations would be achieved. A new authoring mechanism called 'virtual authoring' for multimedia documents was discussed. This can be used to create documents by integrating and customizing segments of existing documents. Virtual authoring provides the framework for the new information layer on top of the existing multimedia documents. Usage of existing documents in developing new documents without replication and without violating copyrights was suggested as a chain for information reuse. Virtual authoring could act as an alternative – as a new mechanism – to link multimedia information sources.

ICDL 2004 SUPPORTERS

Partner Department of Culture, Ministry of Tourism and Culture, Government of India

Associates Department of Information Technology, Ministry of Communications and Information technology Government of India • Commonwealth Educational Media Centre for Asia • Centre for Development of Advanced Computing

Co-Associates Emerald • Goethe-Institut, Max Mueller Bhavan • Italian Institute of Culture • Public Affairs Section of the (Speaker Sponsors) American Embassy • French Information Resource Centre, Embassy of France in India • The British Council, India • Embassy of Austria in India

Function Hosts Blackwell Publishing • THOMSON-DIALOG • Institute of Electrical and Electronics Engineers • Global Information Systems Technology Pvt. Ltd • Murugan Magazine Distributors Pvt. Ltd • Microsoft Corporation (India) Pvt. Ltd

Sponsors Nature Publishing Group • Tausco Book Distributors