Intl. Inform. & Libr. Rev. (2003), **35**, 281–309
doi:10.1016/S1057-2317(03)00014-6

Building Digital Resources: Creating Facilities at INSA

USHA MUJOO-MUNSHI*

Abstract

Today libraries are at a transition phase where twin processes of paper-based environment and changing information-seeking patterns in the electronic/digital environment go hand-in-hand. Hence, all components of the information chain are in a state of flux. The rapid growth in computer and communication technologies have greatly benefited the advanced countries, while the developing countries have not adequately reaped the benefits of such facilities to the desired extent. The application of information technology (IT) in India started on a very modest scale. During the past decade or so several Indian libraries have initiated activities to create, acquire, and provide access to electronic resources. The establishment of networks has had a great impact on libraries and information centers (LICs) in the country, and have further buttressed the IT applications in the LICs to a certain extent. The emergence of the Internet, especially the World Wide Web (www), added a new dimension to information creation and delivery, which also globally triggered digitization programs. Buying access or acquiring digital resources started taking root. The digitization of records (document management) crept in, which attracted librarians and people from other professional backgrounds into records management. This was followed by content management, (currently a popular phrase in this part of the world), also known as digitization. The digitization of documents is now becoming a major activity in libraries and archives. The Indian National Science Academy (INSA) is a premier scientific body engaged in the dissemination of information to the scientific community at large, publishing and promoting scientific endeavors, besides having other multifaceted human welfare-oriented activities. The growing acceptance of digital media has resulted in libraries buying and

* Head, Informatics Center, Indian National Science Academy, Bahadur Shah Zafar Marg, New Delhi, 110 002, India. E-mail: newexp@nda.vsnl.net.in; umunshi@hotmail.com

providing access to Internet resources, acquiring CD-ROM-based data-sets, and providing services for stand alone or networked CD-ROM environments, and digitizing documents. The Academy library facilitates all three. The Academy has initiated several digitization initiatives for content development and management by way of the scanning of publications, image management, and conversion from digital documents to web-enabled resources. The Academy has adopted a three-pronged approach of providing access to digital resources, and acquiring and creating digital resources, for which INSA suitably augments with IT infrastructures and takes initiatives to provide links to requisite data sets for the benefit of its users. INSA developed and provided IT facilities at a modest scale to its users at a time when only a limited few had developed such facilities in the country. The facilities developed at INSA augur well with the initiation of pilot and sponsored projects pertaining to digitization of records and making provision for creating digital resource bases, thereby contributing to the national digital repository on the one hand and providing access and visibility to national resources on the other. The article dwells upon various elements that have contributed to providing services in the changing information seeking patterns of users in the electronic environment, and the building of digital resource bases, while facilitating others to get involved in digital content creation activities. It is hoped that such endeavors shall help in the building up of a national digital knowledge resource base for the country, and INSA would in the process act as a facilitator.

INTRODUCTION

Today libraries are at a transition phase where the twin processes of paper-based environments and changing information-seeking patterns in the electronic/digital environment go hand-in-hand; hence all components of the information chain are in a state of flux. Over the last 50 years or so there has been a radical change in the conduct of routine business in many organizations due to the developments in computer and communication technologies. These technologies have become pervasive in all walks of life. This has buttressed the application of Information Technology (IT) tools and techniques in library and information centers as well. However, it appears that rapid growth in computer and communication technologies have greatly benefited the advanced countries while the developing countries

have not adequately reaped the benefits of such facilities to the desired extent.

The application of IT in India started on a very modest scale. During the past decade or so several libraries in the country have initiated activities to create, acquire, and provide access to electronic resources. The activity started with the development of the computerized catalogues, which paved the way for automated circulation functions with modest facilities. As things developed circulation systems could interact with catalogues, which further paved the way for automating more services with additional features made possible by technological advances in both hardware and software. The issues pertaining to standards and formats were partly solved. Other databases were created, and functions such as cataloguing, acquisitions, serials, circulation, interlibrary loans, financial control, stock management, and user details were all integrated and automated to a certain extent. The electronic communication of information/data from one library to another had its initiation using dial-up connections. Remote database access for providing reference/bibliographic services also began. And, a chain of networks was established nationwide at all levels (local/regional/national) to promote resource sharing. These networks impacted Indian libraries and information centers ("LICs") and have further buttressed the IT applications in the LICs to a certain extent[1].

The emergence of the Internet, especially the World Wide Web (www), added a new dimension to information creation and delivery, which also globally triggered digitization programmes. Buying access or acquiring digital resources also started taking root. By this time, the digitization of records (document management) crept in, which attracted librarians, besides people from other professional backgrounds, into records management. This activity remained in vogue, but the very concept of content or knowledge management was missing because the whole process of flexibility and high precision recall and relevancy were not addressed. The process was more like finding a document or drawing some data. This gave rise to "content management", which is presently a popular phrase in this part of the world, and is more often referred to as digitization. The digitization of documents is now becoming a major activity in libraries and archives.

The libraries and archives that contain some of these valuable resources are either beginning to digitize their documents or are trying to find resources to do so. While a substantial proportion of libraries have automated their basic library functions, particularly retro-conversion of their catalogue records and circulation activities, some provide for automated personalized information services, web-enabled information services, and

[1]Usha, M.M. (Ed). (2002) Information management in new millennium. New Delhi: Allied Publishers, p. 679.

digital content access and acquisition. Only a very few, however, have initiated digitized content creation activity. The initiation of digital content creation is limited to digitization of theses and dissertations, Hindi gazette records, etc., by the libraries. And, individual organizations and publishers are now making full texts of selected journals available on the web.

Mindful of the Government's concern to provide facilities for state-of-the-art libraries, with an emphasis on content creation, there is a definite trend in the move from traditional to electronic/digital practices based upon the changing information-seeking behaviour of users in the electronic environment. Several funding bodies have come forward to promote IT-based information generation, organization, and dissemination activities. Agencies like the former National Information System for Science and Technology (formerly known as NISSAT; now in acronym limbo) Program under Department of Scientific and Industrial Research (DSIR), were initiated with the objective of organizing information support facilities for researchers and academicians. It has been continuously reorienting its programs and activities in tune with changing global scenarios and national efforts of liberalization and globalization of economies. Major efforts of NISSAT have been to establish linkages between information resource developers and users in India and other countries. Its current thrust is on digitization and content creation. In addition, the Ministry of Communication and Information Technology (MIT), the Ministry of Human Resources Development (MHRD), the Council of Scientific and Industrial Research (CSIR), the Defense Research Development Organization (DRDO), and others are also currently funding such initiatives. These agencies have extended grant-in-aid facilities to a number of organizations worth billions of rupees over the last few years to support various programs and projects in the area of information generation, organization, and dissemination. These facilities are and have further buttressed the use of IT tools and products by the libraries and information centers in the country.

The present document attempts to focus on some initiatives taken by the Indian National Science Academy (INSA) pertaining to creating facilities for building digital resources. All initiatives indicated below have been executed by the INSA Library, which was renamed the Informatics Center in 1996, emphasizing the changing catalytic role the center plays in fomenting changes in information retrieval, dissemination, and communication channels in the Academy. However, in the present paper, both the terms "Library" and "Informatics Center" have been used for entities representing the INSA Library in general.

The INSA, established in 1935, is a high-level body of scientists representing all branches of science and technology. INSA is charged with the objectives of promoting science in India, safeguarding the interests of Indian scientists, establishing formal ties with international bodies, and harnessing

scientific knowledge for the cause of humanity and the national welfare. INSA programs are staffed by only 92 regular staff members and a few professionals for discrete projects, serving the large, widely-dispersed Indian science and technology community.

IT Infrastructure, Campus LAN and Internet Facilities at INSA

INSA has suitably augmented its IT infrastructure. Prior to 1994 only a couple of stand alone computers and printers were in place. Since 1995 provision has been made to provide access to computers to all. The present infrastructure stands at over 90 computers (of various makes and models — IBM, Macintosh, and the like), high-end servers, printers, (both networked and others), scanners, CD-writers, and related state-of-the-art equipment for carrying out the work in a changed electronic office environment.

In addition, the CD-Net system was the first of its kind installed in India, in 1995; it will soon be upgraded. The system has 12 nodes located at various points where the science/technology and research/development communities can access these facilities. The Academy began providing access to CD-based information services through its CD-Net system in the LAN environment in the Guest House as far back as 1995, when such facilities were almost non-existent in India. The components of this system are given below:

CD-Net tower
Model-556 M-Q66
80586DX @ 66 MHz
Quad speed 150 m/s SCSI drive
16 MB RAM
32 bit
SVGA color monitor (memory 1 MB)
1·44, 1·2 MB

CD-ROM drive bays 56
No. of drives 28
Ethernet card (32 bit)

660 MB-capacity each
153·6 kB/s data transfer rate
or wide SCSI interface with
300 kB/s data transfer rate
150 ms average

Network server
80586DX @ 66 MHz
16 MB RAM expandable to 64 MB
32 bit
IGB (with provision for second H/D)
SVGA color monitor (memory 1 MB VRAM)
1·44, 1·2 MB

CD network nodes (12 nodes)
(IBM PCs) 80486 CPU @ 66 MHz
32-bit EISA bus
4 MB RAM expandable up to 64 MB
1.2.1.44 MB FDD
320 MB/HDD
6 expansion slots SVGA color monitor (memory 1 MB VRAM)
(with VGA color display card Res.1024 * 768, 256 colors)
101 keys keyboard
2 serial, and 1 parallel port
RTC with battery backup
Ethernet card 32 bit

LAN accessories
UTP (within the buildings)
BNC (between the buildings)
Hubs

Modem 14400/56000 KBPS error correction

Backup network server
80586DX @ 66 MHz
16 MB RAM expandable to 64 MB
32 bit
IGB (with provision for second H/D)
SVGA color monitor (memory 1 MB VRAM)
1·44, 1·2 MB

Network software
(Novell NetWare)

The web server is a fairly a high end Sun Solaris system with the following configuration:

Sun fire 880 server @ 750 MHz.
8 GB memory. 6 × 36 GB.1.0″, 10 000 RPM,
FC-AL disks, DVD, 3 (N+1 redundant)
power supplies and redundant cooling fan trays
17 inch entry color monitor

PGX64 24-bit color frame buffer
Software on CD

Type 6 USB power cord kit

Solaris 8 standard, latest release
English-only media kit

10/100 ethernet adapter

20GB 4 mm DDS-4 internal tape drive

Scholar PAC workshop compiler 6 · 02
Includes unlimited user domain license
for C, C++& HPC

Sunscreen secure net 3 · 1
Unlimited user license for firewall

The contents of other scientific organizations are also expected to be hosted on this server.

The campus LAN at INSA consists of a combination of hubs and switched and routed networks with fiber optics support; enhanced CAT-5 UTP cabling and CISCO Catalyst 2924XL Layer-2 Switches with 1000BASF-LX/LH Long haul GBIC, CISCO 1720 Modular Routers with 10/100 Base T Modular Router w/2 WAN slots; and, 8 M Flash/32 M DRAM. Fig. 1 illustrates the data network solution of INSA.

For the present, 289 access points are provided, distributed over three buildings: The Informatics Center, the Convention Center, and the Jubilee Building. Such coverage is also being extended to the canteen, besides further enhancements in convention facilities and the Guest House. It has been planned to provide cyber cafe-type facilities in the Guest House, thus bolstering communication, information retrieval, and information dissemination. For the present traffic scenario, only 64 kbs has been considered. However, as per the future planning for hosting the contents of other

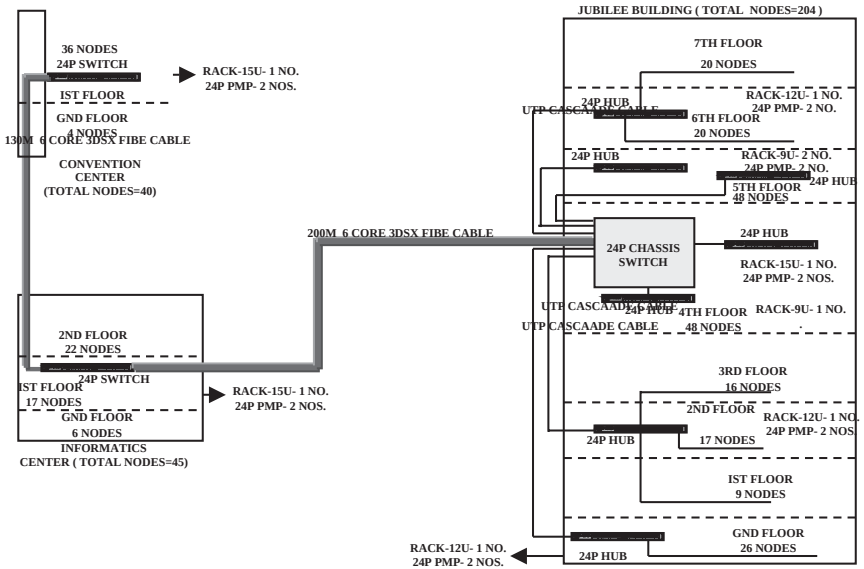

FIGURE 1. Data network solution — INSA.

science and technology organizations, the link shall be suitably upgraded to have better broadband connectivity. This was also done to reduce the cost of investment at the moment, since other related issues were to be addressed on a priority basis.

It may be worthwhile to mention that one of the leading Indian ISPs, known as Videsh Sanchar Nigam Limited, opened up its Internet facilities for the public on a commercial basis in August 1995 and INSA took up PSTN-based connectivity in October 1995. INSA then extended these facilities cost-free not only to the INSA staff and the library's public users, but also to transit visitors in the INSA Guest House, which was the only academic center having such facilities at that time and perhaps the only Guest House of any agency — private or public — in the country to have information access facilities through the CD-Net system in a LAN environment and 24-hour Internet access.

It may be worthwhile to mention here that the Academy caters to the science and technology community in general and its Fellows found throughout India. Since most of the principal users of its services are located far and wide, it became all the more imperative for INSA to make provision for providing services in remote mode using modern tools and techniques. Accordingly, the INSA library constantly endeavors to stay abreast of the information needs of its users in a changing electronic environment, and is making provision for more and more services.

BUILDING A DIGITAL RESOURCE BASE

The INSA library policy, which came into force in 1995, is to make all possible efforts to create, acquire, and access electronic resources.

Creation of records
Efforts have been made to create records of various resources for effective and efficient information, organization, management, and retrieval. Some such efforts are outlined below.

Bibliographic records. The library of INSA has a special collection of books and good resources in the areas of philosophy of science, history of science, science biographies, science and technology policy, science education, science and society, collected works of scientists, emerging interdisciplinary areas, books published by fellows, and reference books on science and technology. However, textbook collection does not come under the purview of the library's mandate. The retro-conversion of 30 000 books was initiated in 1994 on a full scale and has since been completed. The retro-conversion job was undertaken using the CDS/ISIS — software package of UNESCO. The data is also made available on a windows platform using indigenously-developed software consisting of small conversion programs. This is done because all our other modules such as acquisitions, serial control, and circulation are available on this indigenously-developed software. However, CDS/ISIS continues as the INSA library is participating in the regional network systems like the Developing Library Network (DELNET) (earlier Delhi Library Network), and the data needs to be exported to DELNET's union catalogue, for which records are contributed by the participating libraries. The Online Public Access Catalogue (OPAC) of the library is available, for which several nodes have been provided.

Database of serials received in the INSA library. The library has useful literature, as reported by eminent scientists in several dailies in their science and technology columns. It is also presently receiving over 1000 periodicals, a substantial proportion of which is coming on an exchange basis from 368 scientific organizations and academies worldwide. The serial database creation has been initiated using indigenously-developed software and is currently ripe for updating.

Other databases. The other databases include the databases for the personalized information services such as the (1) Selective Dissemination of Information (SDI); (2) Current Awareness Service (CAS) from over 380 journals; and (3) Contents Abstracts and Photocopy Supply (CAPS)

service from over 800 journals, covering eight broad areas of science and technology. The electronic databases for such services have been created using indigenously-developed software from which the information for the registered users of these services is automatically generated once every 20 days. However, the CAS service does require human interface as well in its final phase. The account keeping system for these services has also been automated, from which bills for payment are periodically generated and sent to users.

Database of e-resources acquired in the library. The database of e-resources being acquired in the library is also maintained and available to users. As indicated above, the library has initiated building an e-resource collection, which it is gradually strengthening. The library has compiled a database of its CD-ROM collection of various resources procured by the library, which has over 100 records at present.

Expert database. An expert database consisting of visionaries of Indian Science who are also fellows of the Academy from various areas of science and technology has been created. In addition, the database for the foreign fellows of the Academy is also available for user access. These databases include detailed descriptions of the fellow, giving the name/s, present position, their correspondence address, residence address, contact numbers, e-mail, FAX, area of research, awards received, and other recognition (e.g., membership on important committees), etc.

Other specialized databases

The other specialized full text databases that have been recently generated out of funded projects by the library and are available in digital format. Some such databases are:

- Networking and Communication for IT Jatha.
- Theme pavilion on scientific inputs for promotion of information technology.

These databases are meant for general mass consumption and are being promoted at ITexhibitions in various metropolitan areas, where audiences/visitors to an exhibition at one location may participate in video-conferencing sessions and interact with visitors in other locations.

The databases are full text databases covering areas like advancements in electronics and circuitry, data recording systems, communication lines, voice and optical recognition systems, and other related areas in networking and communication. Easy and interactive access to the contents

and online question/answer sessions for young visitors have also been made.

The National Center for Science and technology Communication (NCSTC), Department of Science and Technology, funded both of these projects.

Science education initiative. Under this program, it was proposed that INSA would have a dedicated website on science education in the country that would also help in networking individuals and organizations through knowledge exchange and experience sharing for the improvement of science education, with a special emphasis on taking science to the masses and encouraging women to consider careers in science. This would focus on issues such as experience sharing, improving rural science education, and disseminating information on the use of experimental materials and inquiry-based learning modules. Based on the information received from participants, strategies would be identified and implemented. As a sequel, a seminar was organized on Trends and Future Initiatives in Science Education, highlighting new initiatives, methods, and models. Also considered were techniques in science education developed by both government and non-government organizations, and individuals. The initiatives taken by the participating institutions that were invited to participate in the two day seminar, May 16–17, 2002, and who presented their initiatives in a variety of posters, science kits, and educational tools, were ultimately digitized and shall be posted on the web for wider dissemination. The material has already been digitized and mounted on CD.

Database of publications received under publication exchange program. The database of publications received under the exchange program is maintained by the library, along with the database of participating publication exchange program agencies. The database gives complete details about what is being sent, received, the year of initiation of exchange, traces of changes in exchange scenarios, and major events and changes. The Academy has publication exchange relations with 368 scientific organizations/academies worldwide, with over 700 participating journals in the exchange involving the Academy's four research journals, yearbook, and other special publications.

Database of video collection of eminent scientists. INSA has prepared video strips of various eminent scientists highlighting their research and development efforts. The activity was initiated recently and the collection is maintained by the library, with easy access by users. This collection is being continually updated with the passage of time.

TABLE I

Cassell concise dictionary
Classics of Progress: Selected Papers and the General Index of the Progress of Theoretical Physics
Digital Multimedia Reference
E-Commerce: The EDI Way IT Encyclopedia.Com
Encyclopedia of Networking
Encyclopedia Britannica
Eyewitness Encyclopedia of Nature 2 · 0
Eyewitness Encyclopedia of Science 2 · 0
Family Health Encyclopedia with Interactive Diagnostic Charts to Help Identify Symptoms, Disorders and Emergencies
International Who's Who 2000
Manorama Knowledge Adventure 2001
McGraw Hill Multimedia Encyclopedia of Science and Technology
Multimedia in Action
National Geographic 1889–1999
Science Citation Index 1980–June 2002 (being received regularly on quarterly basis with yearly updates)
World Book Multimedia Encyclopedia
World Wide Web Yellow Pages

Digital resources acquisition

The aim of acquiring digital collections is primarily to meet information needs as quickly and efficiently as possible. The library's collection development policy emphasizes the acquisition of e-resources and has established a separate budget entry for this. The Academy started from scratch to build a digital collection. As a first step it provides for building up a digital collection consisting of (a) secondary resources in digital format only; (b) back-runs of journals on CD-ROM or online; (c) other resources and e-books. The library has since initiated collection development of e-resources in a phased manner. The library now acquires secondary sources only in electronic format. The back-runs of some of the journals that are available in CD-ROM format are also being acquired. The hard copy, together with the soft copy on CDs for books is also being purchased and maintained in the library and users are encouraged to make use of such collections. The following are a few selected titles (Table I) acquired by the library, and such collection is being strengthened on a continuous basis. Besides these, there are over 100 titles for books and other resources available in the library in CD-ROM format for access by users to the library.

Access to digital resources

The library has been rendering service to users by providing access to electronic versions of journals that are being received by the library in hard

copy format. Moreover, the science and technology community also has ease of access to e-journals available in the public domain. A few Indian journals, which are in the public domain, are also being used for this purpose.

Value-added e-services. Repackaged information services from e-resources based on web resources in the public domain and CD-ROM-based resources are also provided to our remotely located users electronically. Some such services also need human computer interfaces at certain stages.

CD-ROM-based search services through CD-net system. The advent of CD-ROM-based technology in the 1980s brought about a revolution in acquiring affordable e-versions of several bibliographic databases on CD-ROM. For instance, accessing DIALOG databases in India a decade ago would have cost $3 (at that time, $1= Rupees 30–35) as a minimum access charge for 3 min or any part thereof. Only researchers with plentiful institutional resources could have afforded the service. Hence, the research and development community in developing countries could barely afford to access databases unless they were well-financed. Multifaceted CD-ROM-based databases (bibliographic or full text) became very popular in developing countries. Their affordability, portability, and durability (e.g., withstanding extreme environmental conditions), and usability made them popular. Soon online access to databases was by and large replaced by CD-ROM-based databases wherever possible.

The CD-ROM-based services were initiated in the INSA library in 1995 when a CD-ROM-Net system was acquired and installed, the first of its kind in India. The collection development policy for databases generally does not favor acquisition of subject-oriented databases, since such databases are available in various neighboring libraries of INSA. However, the mandate of the library provides for acquisition of general types of e-databases generally not procured by other libraries. An initial subscription was to the Science Citation Index (SCI) in CD-ROM format. The complete file from 1980 to date has been procured. The database is available at only two locations in the metropolitan New Delhi region.

Access to this database is provided to all those who do research and development in scientometrics/bibliometrics, and to others for general reference search purposes. Users have been provided 12 nodes to access this facility, with nodes also made available in the INSA Guest House for access by transit guests.

Scientific output evaluation service. The SCI is extensively used for cytoanalytical studies in various areas of science and technology for mapping science,

identifying trends in scientific research, identifying institutions of excellence, and evaluating scientific output in academic competitions. Both individuals and organizations are requesting such information services. However, users desirous of conducting such searches themselves are also accommodated by the system. This service has become very popular.

The other databases that run on this system are the databases indicated in this document in Table I.

Provision for access to J-Gate. J-Gate is an Internet gateway and portal set up nearly 2-years ago by Informatics (India) Ltd. It offers affordable access to global electronic journal literature and provides seamless access to journal articles through the massive database interface of 10 000+ e-journals. J-Gate as a portal covers content from thousands of e-journals published worldwide.

Modalities have been developed with J-Gate publishers M/s. Informatics to facilitate J-Gate access to INSA Fellows both in India and overseas. The fellows can access the facilities provided by J-Gate. The password rights have been extended to the fellows of the Academy to e-walk through J-Gate. Currently J-Gate offers two types of products/services:

- J-Gate Portal
- J-Gate Customized Services

J-Gate Portal offers:

- Directory of e-journals-10 000+ journals listed with links to journal/publishers sites.[2]
- Table of Contents (TOC)- 9000 + journals
- Database — A comprehensive searchable database — 10 00 000 + articles added every year.

J-Gate initially has provided J walk through its package to INSA fellowship for a limited period that shall be extended further upon mutually agreed terms and conditions.

J-Gate customized services offer:

- J-Gate Custom Content (JCC) — Local Intranet solutions for libraries for journal subscriptions.
- J-Gate Custom Content for Consortia (JCCC) — JCCC is extended to an homogeneous group of libraries for sharing "subscribed" journal resources.

[2]J-Gate — The e-journal Gateway. In electronic format: http://www.j-gate.informindia.co.in

JCCC: Such a facility is already operational. However, the modalities are also being developed for JCCC since this proposal is economically more viable. Of course the drawback in this is that since all the participating libraries in the vicinity of this consortia have to come to a common consensus. It is proposed that INSA would act as a nodal point for such consortia and would facilitate electronic access to other libraries.

Buying and providing access to online journals. The INSA advisory committee has decided to strengthen the e-collection for remote access and has emphasized the consortium approach. Since INSA is unique in that its users are found nationwide, it has recently been decided to strengthen access to e-journals rather than subscribing to new titles. Efforts are on in this direction to buy e-access to both existing subscribed journal titles and also new titles that are likely to be acquired by the library. In addition, INSA also shares a unique position among science and technology agencies due to its being a premier scientific body catering to the needs of the science and technology community. INSA also promotes scientific endeavors by various means and mechanisms. Therefore, almost all the resource centers of science and technology bodies share their resources and serve INSA by providing direct or indirect remote access to their e-resources for INSAs users. The INSA is making all possible efforts to provide access to users and to remove bottlenecks in information access. INSA is also bridging the gap between information haves and have nots by facilitating remote access to resources of various leading institutions. Since Internet connectivity is by and large available at all places (having broadband or PSTN connectivity), such facilities are being extended by sending information electronically or making provision for them to access e-resources, thereby bridging minds over miles.

Current access to over 1000 + journals is available in the public domain. Access is also provided to some Indian journals that are available online at http://www.infolibrarian.com/ejls.htm.[3] Thus, digital resources are available to fulfill user requirements effectively, efficiently, and in a timely manner.

Recently, participating in local/regional/national networks has proved to be the only viable solution in situations having static library budgets coupled with rising serial costs. The statistics show that over the decade 1986–96, journal costs have risen by an average of 148% compared to about 44% in the consumer price index.[4] What is more, libraries are spending approximately 120% more than they were in 1986 and yet get 7% fewer titles. The Indian scenario depicts that roughly around 15% of

[3]Infolibrarian.com. In electronic format: http://www.infolibrarian.com/ejls.htm

[4]Malakoff, D. (1998) *Science* **282**, 853.

the foreign subscriptions were dropped.[5] Though many of these journals are available online, many institutions are nonetheless not in a position (economically) to buy access. However, some consortia like the CSIR have been successful. INSA is also moving in this direction, playing a proactive role and strengthening the consortia approach for online access to journals, thereby enabling the research and development community to have timely access to a wider resource base.

Provision of bibliographic/abstract/full article services. The INSA library has also made provision for compiling bibliographies from e-resources and mailing them electronically to end-users. In many cases, bibliographies are complemented with abstracts as well, while full article services are generally restricted to fellows of the Academy only.

Access to DELNET databases. The library is actively participating in regional/national networks. INSA has been participating in the Developing Library Network (earlier Delhi Library Network) activities from almost DELNET's inception in 1988. As a participating institution, INSA is contributing its data in e-form for DELNET's databases on books, journals, and other similar resources. INSA also participates in document lending services in a very proactive manner. INSA is facilitating access to the following DELNET databases and services (http://delnet.nic.in)[6] and is also allowing access to resource base of 460 libraries by INSA (Table II).

Digital Initiatives: Conversion of Existing Print Media into Digital Format

The digitization of valuable original materials is often undertaken with the dual goals of both improved access and enhanced preservation. The most important parameter of a digital library is the digital collection — either its acquisition, or provision to its access, or creation for present use and for archival purposes. The ease of creation, use, and access has led libraries to start initiatives pertaining to various aspects (access, acquisition, creation) of digital resources.

Content developers, be they academics (at universities or colleges), publishers, or information generators and disseminators, have developed strong interests in electronic publishing. For example, many universities in the West are producing their publications in electronic format and hosting

[5]Personal Communication.

[6]Delnet–Developing Library Network: In electronic format: http://delnet.ni.in

TABLE II

Delnet databases	No. of records
Union Catalogue of Books — CCF	16 65 634
Union Catalogue of Books — MARC	44 132
Union List of Current Periodicals	20 777
Union Catalogue of Periodicals	19 228
Databases of Periodical Articles	4 22 217
CD-ROM Database	1 669
Union List of Video Recordings	3 376
Union List of Sound Recordings	748
Database of Urdu Manuscripts	210
Databases of Theses and Dissertations	32 553
Indian Specialists Databases : a Who's Who	2 000
Devinsa Database	20 000

them online for wide, global, and 24 hour access on the one hand, and for preservation purposes on the other. Important endeavors in this field have been undertaken by the Digital Library Federation of the US, including the University of Pennsylvania Project for Archiving and Preservation of Electronic Journals, Project Harvest of the Cornell University Library, the Digital Preservation Collaboration between Yale University and Elsevier Science, the Harvard University Library Electronic Journals Archiving Project, and the MIT Dynamic e-journal project. On the publishing side, thousands of journals have been hosted online. It appears that many journals are only coming in the digital format with no print versions and the trend seems to be accelerating.

The digitization of valuable original materials is often undertaken with the dual goals of both improved access and enhanced preservation. Whatever strategies or techniques are adopted to preserve digital data, they will only be successful if the data is fully documented throughout its lifecycle. A number of bodies and projects are currently grappling with the issues of preservation of metadata and publishing detailed recommendations as to its implementation. For instance, the CEDARS project (www.leeds.ac.uk/cedars/) has produced a specification for preservation of metadata that is required to support meaningful access to the archived digital content and includes descriptive, administrative, legal, and technical information.[7] CEDARS closely follows the reference model for an Open Archival Information System (OAIS) in its metadata schema, and has implemented

[7]Lupovici, C. & J. Masanes, (2000) *Metadata for long term preservation of electronic publications*. The Hague: Koninklijke Bibliotheek.

the schema in XML. The Networked European Deposit Library (NEDLIB) is also using the OAIS reference model. The PANDORA (Preserving and Accessing Networked Documentary Resources of Australia) Project at the National Library of Australia has developed a logical data model which integrates all the metadata for digital objects consistently (www.nla.gov/pandora/ldmv2.html).[8]

The costs of digital preservation are high. But the cost of not preserving our heritage will be even higher in terms of lost data and lost history, and there is no time to waste in putting strategies and practices in place to preserve key materials. Many individual projects and institutions are working hard on this problem, and there are also national and international initiatives collaborating to find long-term solutions.

We are in the midst of a second revolution in scholarly communication, viz., the era of electronic publishing and the advent of e-journals. Web publishing has broken all the barriers between the creators, users, and providers of scientific information. Today there is turmoil in the scholarly communication system and there is accordingly an opportunity for Indian journal publishers to break free of the existing system by producing electronic versions of journals and hosting them online for global access. There are only a few efforts in this direction such as the IndMed database (by the Indian MEDLARS Center, New Delhi), was the first web-based Indian biomedical database covering titles and abstracts of about 40 Indian journals such as *Indian Pediatrics*, *Down To Earth*, *Bombay Hospital Journal*, *Indian Heart Journal*, *Indian Journal of Community Medicine*. The dearth of similar initiatives is due to:

- fear of long-term visibility;
- no encouragement of previous experiences;
- lack of experience;
- loss of independence;
- duplication of work; and,
- perhaps fear of loss of the subscription base, itself a questionable prospect.

In India, the IT Task Force set up by the Government of India has recommended the need for content creation in digital formats, at all levels. As indicated above, there have been a few endeavors made by both the institutions and publishers in this direction. With the thrust and impetus coming from various sectors, the e-culture is gradually rooting itself in India and many institutions are coming forward to initiate digital content creation activity. However, it may be mentioned that no other library in India has

[8]In electronic format: www.nla.gov/pandora/ldmv2.html

initiated digitization of its organization's journals with the same alacrity as the INSA library.

Many libraries have gained experience with aspects of digital resource management. Some have public access to the Internet, others have substantial numbers of networked CD-ROMs, while a few have scanning capabilities. INSA's Informatics Center facilitates all three of these activities. INSA has also undertaken several digital initiatives for content development and management by way of scanning publications, and managing conversion from digital to web-enabled resources.

INSA's initiatives

To meet its objectives, INSA has developed various specialties such as the publication of scientific research. INSA is publishing both regular and special publications, and has produced over 200 special publications. In addition, INSA publishes four science and technology journals and two other regular serials. For the past several years these publications have been available digitally. The INSA library also has a rich resource base pertaining to fellowship records, containing highly valuable information unlikely to be available elsewhere in the country.

INSA has been deeply involved with the development of contents regarding science and technology publications in association with other academies. In this direction, INSA has taken several initiatives in content creation using the digital technology on a modest scale. Such initiatives include:

- Digitization of Library Records;
- Indian Journal Server Initiative;
- Initiatives in Pipeline;
- Others.

Digitization of library records. The Library and Informatics committee's decision to digitize library records was made in 1997. As a first step towards that goal, it was decided to digitize journals in brittle condition. These vulnerable pieces included popular science titles, fellowship records, and older INSA journals and council minutes. The initial phase began with the digitization of fellowship records.

Digitization of fellowship records. This project was taken up basically to preserve the important information content of fellowship records available in print form in files. Since most of these files dates back to the 1930s, the quality of the paper has really deteriorated and was difficult to maintain and preserve properly.

Secondly, in this electronic environment, it is more effective and efficient to access such documents without infringing/compromising on the original format and details.

The academy in 1999 initiated this project to digitize fellowship records (the first project of its kind by an Indian library) consisting of important personal, professional, and academic details about fellows of the Academy, together with their research and development achievements. It may be mentioned here that the Academy is engaged in promotion and excellence of science in the country, and also of projecting the endeavors of leading science and technology scholars overseas. Since digital technology at that time was not as advanced as is it is now, the job of digitizing fellowship records was initiated. For this particular job, it was decided to capture the data and create an image-base for the full contents of these fellowship records. The DataScan software is being used for creation, organization, and retrieval of information. The software was customized according to INSA's requirements.

For the data manipulation and effective access to the image-base, a database was created containing keys to the full text. This database has 16 key search parameters for search and retrieval purposes (Table III). Nine of these fields allow the use of keywords. This database also allows the addition of annotations. The image-base in this case has been also made available on the CD-ROMs. While it is very important to keep the index that is stored in the database as indicated above for all documents captured digitally for archival purposes on the hard disk), the contents can be mounted on CD-ROMS as well. In case all the index is not available on the hard drive, it may not be possible for the system to address the user's query, as the indexing terms that act as an interface between the index database and the actual document stored in the

TABLE III

Fellowship code number
Name
Date of birth
City
Sectional committee
Year of election
Photograph
Subject/research area
Awards (annotations) (more details are given at "Research Profile")
Academic record
List of publications
Brief biographical sketch (annotation: for detailed research contribution go to keyword "Research Profile")
Research profile (annotation: for complete list of publications go to list of publications)
Updated information
Personal information
Keywords

image-base is not readily accessible; hence, it gives rise to a missing link. Therefore, it is important to keep an entire index database on the hard drive while an image-base containing actual full text documents can be mounted on CDs. Since the data is massive and cannot be contained on one CD-ROM, in such situations the interface between the index and the document needs to be provided so that while users access the information they learn on which CD-ROM the data is stored. This happens in situations when the user is looking for information on a specific item; he/she receives the information about the availability of 'X' number of documents that matches his/her request. Once such information is displayed on the screen, the user can access the full contents of such documents. In case the desired document is not available on the hard disk drive, but is available on the CD-ROM, the system alerts the user to the document's location, i.e., the CD-ROM containing the full data. It may, however, also be stressed here that the system demands that each mounted disk has to be provided with unique numbering. This unique numbering also forms part of the index. The project has since been completed and the data mounted on CD-ROM; however, the index is on the hard disk for wider access. This database made available on the Intranet may also be put partially on the Internet in the future. Efforts have also been made to provide CD-ROMs with limited search capabilities to Fellows of the Academy.

Indian journal server initiative. The Indian National Science Academy is a premier scientific body representing all branches of science and technology and always strives to promote and disseminate scientific knowledge globally, effectively, and efficiently. In addition, the Academy always tries to tap the best personnel and resources to foster its activities. The academy recognizes the importance and impact of electronic publishing on science. International publishers like the ICSU Press have been deliberating on issues of electronic publishing in science, particularly from the perspective of practicing scientists worldwide, with benefits for the scientific information chain, and thus more progress for science. The ICSU Press has issued a call regarding the electronic publishing of Indian science and technology journals. The project basically would encourage developing nations to use electronic means to gain access to western literature and also to put their journals in electronic format so as to promote usability (and visibility) of their publications[9] INSA representatives have been involved with the ICSU Press from the very beginning, promoting such activities at the national level also. The

[9]Margaret, B. (1999) E-mail to INSA dated June 17, containing a message from Sir Roger Elliott, Chairman, ICSU Committee on the Dissemination of Scientific Information (usually known as ICSU Press).

academy has also joined hands with another entity: The Indian Academy of Sciences, Bangalore. These two academies initiated the hosting of science and technology journals on a national level on the pattern of Latindex, which provides information about Latin American Journals, and mounted a number of African journals online (www.oneworld.org/inasp).

Building a digital resource-base of science and technology journals and hosting them online. The two leading scientific academies in India have a 60-year record of consistent publication of journals in India. It was therefore natural and appropriate that the two academies combine their efforts to electronically publish journals. As a result, a Joint Academy's Committee for Electronic Publishing (JACEP) was established to develop plans and programs for the two academies to cooperate in putting their research journals on the World Wide Web on Indian servers. As part of the recommendations of this joint venture, each Academy is to set up a web server for hosting journal contents online. The two servers will mirror each other. The infrastructure for high-availability web servers was to have high-bandwidth connectivity. And, a third mirror server would be installed at the Supercomputer Education and Research Center (SERC) at the Indian Institute of Science (IISc.), Bangalore.

Funding for the project and training of the staff were additional critical issues. The infrastructure has since been procured, and staff have been trained in various facets of electronic publishing. The Indian Academy of Sciences, Bangalore, has made some headway and now has the full content of most of its journals online. The Indian Academy has made available its 11 journals online (http://www.ias.ac.in/journals.html),[10] as per the list given in Table IV covering back issues from 1999. However, issues prior to 1998–99 are not yet available on-line.

The following paragraphs provide an overview of INSA's initiatives in this regard, pertaining to the project of building digital resources for Indian science and technology journals and hosting them online:

(1) Objectives. The present project has been initiated by INSA with the following objectives:

- To nurture a national digital resource base with global visibility and accessibility;
- To produce electronic versions of Indian science and technology journals and to make them available on a web server;
- To support research activities by providing global 24-hour access to the full texts of scientific information.

[10] Indian Academy of Sciences Journals: In electronic format: http://www.ias.ac.in/journals.html

TABLE IV

	Name of journal	No. of issues a year	Volume number in 2002	E-mail address
1	Proceedings — Chemical Sciences	6	114	chemsci@ias.ernet.in
2	Proceedings — Mathematical Sciences	4	112	mathsci@ias.ernet.in
3	Proceedings — Earth and Planetary Sciences	4	111	epsci@ias.ernet.in
4	Sadhana (Academy proceedings in engineering sciences)	6	27	sadhana@ias.ernet.in
5	Pramana — Journal of Physics	12	58, 59	pramana@ias.ernet.in
6	Journal of Biosciences	4	27	jbiosci@ias.ernet.in
7	Bulletin of Materials Science	6	25	matersci@ias.ernet.in
8	Journal of Astrophysics and Astronomy	4	23	jaa@ias.ernet.in
9	Journal of Genetics	3	81	jgenet@ias.ernet.in
10	Resonance — Journal of Science Education	12	7	resonanc@ias.ernet.in
11	Current Science (published by the Current Science Association in collaboration with the Academy)	24	82, 83	currsci@ias.ernet.in

TABLE V

	Name of Journal	Year started	No. of issues a year	Current volume
1	Indian Journal of Pure & Applied Mathematics	1970	12	33
2	Proceedings of INSA A — Physical Sciences	1959	6	68
3	Proceedings of INSA B — Biological Sciences	1959	6	68
4	Indian Journal of History of Science	1966	4	36

(2) INSA publications. The project involves digitization of all the back volumes of the following science and technology journals published by INSA (Table V).

(3) Training. A training program providing exposure to the INSA staff in the area of electronic publishing was organized at the Indian Institute of Sciences (IISc), Bangalore, coordinated by SERC, Bangalore, along with IAS, Bangalore. From March 13th to the 15th, 2002, several staffers were again sent to IISc, Bangalore, for further training.

(4) Converting from print to digital. The conversion from print to digital format encompasses a whole set of activities. For existing copies of journals all the

TABLE VI

Author/s
Journal
Title
Volume
Issue
Initial pages
Year
Article title
Keywords (7)

data since the inception of their publications have been processed for conversion to digital format. These journals are available for online access (http://www.insa.ac.in) with intensive search parameters. Presently, the Portable Document Format ("PDF") is generated from hard copy after scanning that uses fairly versatile scanning software. For the search, the metadata or structure is created which includes the bibliographic details about the article, together with several keywords and other search parameters for efficient and effective searching. For instance, the search parameters include searching by the fields indicated in Table VI.

(a) Organization of data. Organization of images and corresponding metadata into a Database Management System (DBMS) is done using Oracle and also on MySQL because the co-servers make provisions for MySQL and Oracle.

(b) Software. The content management system software being used in this case is DXWarehouse-Archive Enterprise Server (developed using Open Standards like EJB, Java Servelets, and JSP), having the following components:

- Media Repository Manager (includes storage and management of all files required for DX Reader);
- DPS XML Index (database index used by the DPS Search Engine);
- DPS Connection Interface Component (required for connection to the DXReader Interface);
- DPS Remote Administrator (administrative functions like data management, file management, and user access management);
- DPS Inter-server Communication and Query Module;
- Unlimited user connection license.

(c) Retrieval interface (web-based browse and search capability). The PDF has been optimized for online delivery. Though XML is the current standard in web markup languages, such provision would be incorporated in future volumes. To facilitate searching online journals in a highly interactive way and providing effective interfaces between the user and the system, so as to facilitate efficient and effective global searching and retrieving, is of utmost importance. The design and development of an effective search strategy is possible with the 14 parameters (seven fields and seven keywords) for effective handling of data manipulation, suitably addressing the users' queries and fulfilling user needs and requirements. Undergirding this effective searching is data organization of images and corresponding metadata into Database Management Systems (DBMS) such as Oracle and MySQL. In the present project Oracle would be preferred, since the Sun Solaris system (the system procured by INSA) fully supports this RDBMS. However, provision has also been made for MySQL as a co-server. For wider dissemination of the journal contents, it is important to mount the data on the CD-ROMs as well. Therefore, provision has also been made to mount the data on CD-ROMs as a back up as well as for facilitating electronic searching for those people who do not have access to Internet and/or do not have adequate connectivity facilities. The CD-ROM version will have adequate capabilities for information searching and retrieval. The schematic representation of a complete workflow from capturing data to converting to PDF and eventually to XML is depicted in the flow diagram given in Fig. 2.

(d) Outsourcing of the job. Retro-conversion: Since this is a massive job and the Academy does not have so much infrastructure in place, outsourcing was the only viable solution for the initial digitizing of the entire contents from Academy journals. The job included conversion of journals from print to digital format, metadata creation, organization of data, retrieval interfaces (web-based browsing and search facility), hosting of journals online on a public server, and maintenance of the site until the procedure stabilizes.

INSA would eventually take up the entire activity of processing journals for online publishing by its own staff in-house. However, it will require help from the outside agency, and the processing of current journals which are "born digital' for online hosting would be carried on simultaneously.

(e) Financial support. The project has been funded by the National Information System for Science and Technology (NISSAT), of the Department of Scientific and Industrial Research, Government of India.

(f) Expected output. Digitization of all the back volumes (starting from 2002 and going backwards) of all the Science and Technology journals published by INSA.

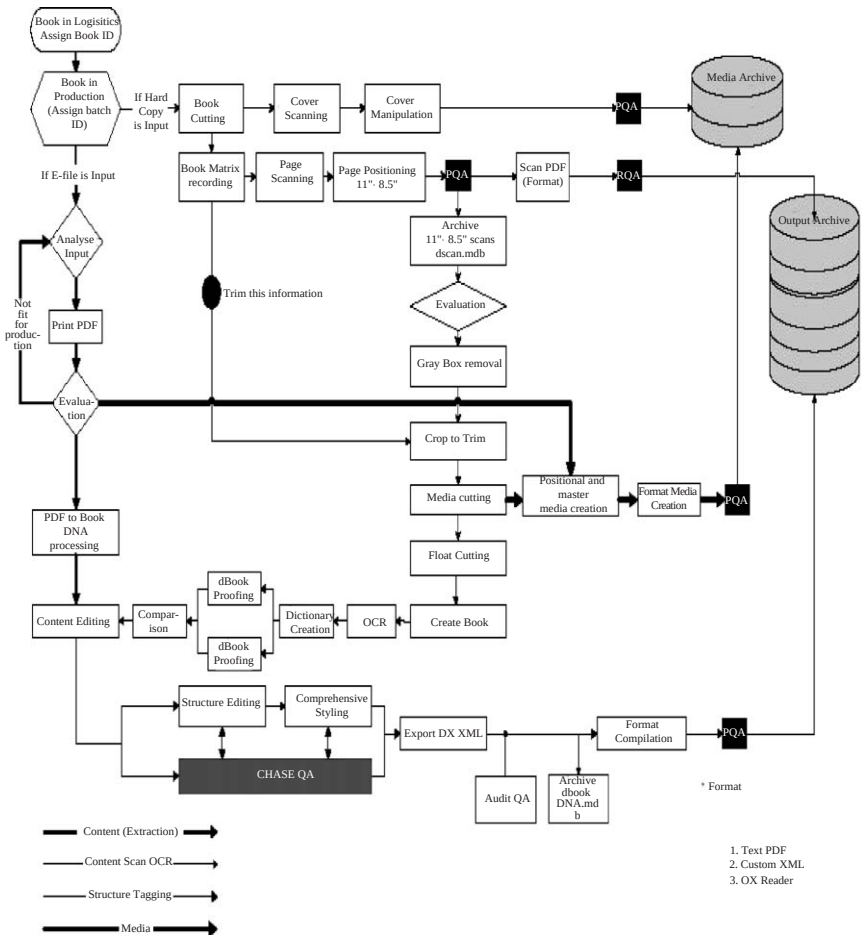

FIGURE 2. Process flow for INSA journals online.

(g) Present status. The project is functioning and is progressing; The contents have been uploaded for global access and are available at: http://www.insa.ac.in.

Digital initiatives in pipeline

Digitization shall soon be undertaken for the following projects:

Biographical memoirs. INSA conceived the idea of producing the Biographical Memoirs as a means of providing an authentic account of the inspiring lives and achievements of the past Fellows of the Academy. These Fellows had

specialized in different spheres of science and played significant roles in advancing the frontiers of their chosen fields. There are at present 22 such volumes.

Fellows of the Indian National Science Academy (1935–93): Diamond Jubilee Compendium. The two-volume publication is a valuable resource giving brief biographical sketches (including photographs) of 1137 distinguished scientists admitted to the Fellowship of the Academy from its inauguration in 1935, up to 1993. These publications are being produced at regular intervals. Efforts shall be made to put forthcoming volumes on the web.

Books and monographs. The Indian National Science Academy has been producing books and monographs on areas of topical interest from the past six decades. The Academy has published a fine selection of books and monographs, some very highly rated and in great demand by the science and technology community. As part of its content creation programme, INSA is going to digitize books and monographs brought out by the Academy during its over 60 plus years of existence.

Creating and hosting science and technology publications of other organizations. As indicated earlier, the Indian Journal Initiative applies to INSA journals as prototypes. However, it is proposed that INSA would subsequently provide a platform to other organizations for hosting their journals and provide requisite help in content creation.

Science publications in developing countries lack visibility and accessibility; the results are poor quality and little credibility. The Indian Science Journals that meet international standards nonetheless face problems of poor visibility. The solution is e-publishing. INSA is proposing to address these issues by selecting quality journals in the areas of science and technology from various Indian scientific academies/ institutions and providing technical assistance, server space and training, etc. It is hoped that such endeavors shall help in the building up of a national digital knowledge resource base for all of India.

Dissertations and thesis. To build up a repository of the dissertations and theses collections, some universities have been approached to undertake digitization of such collections. The requisite activities to this end have begun.

Other initiatives

Efforts are underway for creating state-of-the-art facilities for training professionals in the use of emerging technologies. The funding agencies are being tapped for providing financial inputs. A similar type of Internet

School that is located at the Indian Institute of Science, Bangalore (in the Southern Part of India), shall be created on the INSA campus. Presently, training programs are conducted in collaboration with other organizations on various aspects of information technology for information organizations, with special emphases on processing and dissemination of digital content creation. The response to these programs is quite encouraging. It is with this objective that the initiatives are being taken to suitably augment the IT lab on a regular basis and train staff and prospective information managers, indeed all who are keen to get exposure to the new information technology tools and techniques.

Conclusion

The INSA Informatics Center has endeavored to provide for the building of e-resource collections and for access to the entire science and technology community. The Center has intensified its digital archive activities by initiating various projects, made possible by management's support and by funding agencies that have readily considered and accepted INSA's project proposals. The Center is further promoting content creation activities by collaborating with other organizations and providing a platform to organizations keen on participating in such activities.

The preservation of India's heritage literature, regardless of format, is critical to civilized society. Thus, mindful of the many strategic factors and costs involved, preservation is an area where there are many uncertainties. Issues are hotly debated and different strategies have passionate adherents or opponents. The costs of infrastructure and information provision are particularly heavy for developing countries. The Government of India, through its various agencies like DST, CSIR, ICMR, and DRDO, can support such e-publishing ventures through one time grants and technical support. The dividends of enhancing the global visibility of Indian science should be well worth the effort. And, librarians find themselves expected to manage increasingly technical projects to achieve their goals of delivering valuable information to their ever-growing user base. It is a challenge that must be matched with practical skills. The key is good project planning and risk management for success in a project-based information technology environment. The range of skills needed is expanding, and there is a shortage of experienced staff for digital library and digitization projects. Therefore, there is a need to nurture talent and transferable skills to ensure that skills are developed in a supportive environment with sufficient training opportunities made available in a holistic way. There needs to be a cohesive approach for taking up digitization work at India's local, regional, and national levels. It is important to know and share the experiences of those who

have already undertaken such activities in their organizations. This will be of immense help for determining particulars before initiating a project. This will also help overcome the bottlenecks that agencies have encountered during and before the execution of the digitization projects. Thus, we avoid that all too common pratfall of re-inventing the wheel.