

Digital Archival Strategies for Library Consortia : The IIM Consortium Experience

M.G. Sreekumar*

(on behalf of the Librarians of all the IIMs)**

A library's digital information collection development objective is not far from its original mandate, i.e., to select, acquire, manage, preserve, and finally archive those digital / electronic materials, which will support the current curricular and research needs of the parent institution or organization to which it is attached. There is a paradigm shift, both from the approaches of the users as well as in their access facilities. In most of our endowed campuses and research centers, the user community have access to digital information using personal computers in their offices, libraries, computer centers, Net kiosks etc. Consequently, the Libraries provide (or rather forced to provide) access to a wide variety of materials using an equally wide variety of access mediums to meet the user needs. In the process the library not only give seamless access to the users a wide variety of content categories, but also accumulates a vast treasure of digital content spanning many formats and standards – proprietary as well as open. Unless state-of-art strategies are adopted from time to time the library face problems in many ways. This paper narrates the possibilities and strategies towards creating and operating a depository, digital, central archive of the IIM Consortium resources, in Internet compatible format, and make it available to all the IIMs via the World Wide Web, 24 hours a day, 7 days a week. The archive is envisaged as a robust backup of the vast digital wealth of the IIM Consortium and in addition, a central repository playing the role of an intellectual network nerve centre.

Electronic Information Selection Criteria

Libraries today buy licenses for an ever-increasing number of information resources from a range of different publishers and providers, and use a diverse set of technologies for information delivery [1]. In addition, a wealth of relevant resources are freely available on the Web for libraries to incorporate into their e-collections and to make them readily available to their users. Materials may be in print and/or electronic form; formally and/or informally published; and stored locally, for access via an institution's intranet, or remotely and accessible via the Internet. A number of services are outside the library's control, but nonetheless libraries want to integrate their resources, presenting the information from any particular source within the context of the complete collection. Searching across repositories is only part of the solution. While not all subscriptions lend themselves to electronic delivery, electronic subscriptions offer a great potential for increased value to the entire organization. The move from atoms to bits complicates the jobs of information professionals, but the benefits - competitive advantage, access to information by a wide spectrum of the users, can be tremendous.

Like print materials, digital titles added to the collection need to match the needs of the clientele assuring appropriate scope, content, depth, and quality. The materials selected should be affordable, the content must be timely, bibliographically accessible and in the appropriate language, etc. It is also presumed that there are no technical reasons why the library cannot provide access, e.g., doesn't use a proprietary browser, permits printing, etc., and that their use by library user and librarians will not require an inordinate amount of training.

*Librarian, IIM Kozhikode

**This paper is dedicated to Late Sh. Ashok Jambhekar, Librarian IIMA, who was the guiding force behind the success of the IIM Library Consortium.

Electronic Journals : Direct, Second Party, and Aggregators

Electronic journals are the digital equivalent of their print counterparts. Whereas databases may contain only some full text or links to full text, electronic journals are literally the entire journal. For a good catalog of electronic journals currently available, see the World Wide Web Virtual Library: Electronic Journals [2]. An electronic journal is the whole journal, and libraries can subscribe to electronic journals from publishers or through a second party (e.g. vendors) just like they subscribe to print journals. This distinction between publishers and second parties is an important one, as these are the two ways libraries get electronic journals. Some companies create collections of entire journals and sell access to these collections.

Second party electronic journal databases are different from other databases because in a second party electronic journal database, the entire journal is collected, as opposed to a full text database like "Academic Search Elite", where all, some, or none of any given journal may be included. Examples of second party electronic journal databases include:

1. JSTOR [3], which specializes in making back issues of a wide variety of humanities and social science journals available electronically;
2. Ingenta [4], the most appreciated journal portal, facilitates users to read, print, and email articles from 6000+ electronic journals from 260 publisher partners. Ingenta is the world's largest website for the search and delivery of research articles and it offers access to article summaries from over 27,000 publications linked to the full text of thousands of online publications. For libraries and information professionals, Ingenta provides an extensive range of free and customized online services, from enabling authenticated campus-wide access to subscribed full text articles to creating branded library gateways incorporating management tools and deposit account facilities;
3. Project MUSE [5], which provides electronic access to journals in the fields of literature and criticism, history, the visual and performing arts, cultural studies, education, political science, gender studies, and many others;
4. Emerald Library [6], which contains over 140 electronic journal titles with full text articles published since 1994;
5. Catch Word [7], which provides access to hundreds of journals from approximately 44 publishers.

In India, J-Gate of M/s. Informatics Ltd. [8], is the best example for an E-journal gateway. The mission of J-Gate is to provide an e-content and e-commerce gateway for academic, research and corporate libraries in India to subscribe, access and manage their electronic journals. The vision for J-Gate is to act as:

- An electronic aggregator for several thousand scientific journals
- A common access and search interface for all the journals a library subscribes to
- An insured and dependable journal archive source for the libraries
- A large database for journal literature available, possibly free, to all Indian scientists, scholars, students and professionals for browsing table-of-content and abstracts

- An online e-journal subscription agent
- An online host for Indian journals

Aggregators use to buy the copyright of the journals in bulk and provide access to the journals through a single interface. Examples of aggregators include EBSCO Online [9] and PROQUEST [10]. EBSCO's Business Source Premier and Academic Search Elite are very popular services which attempt to provide management tools for full-text electronic journals. PROQUEST's ABI/Inform is yet another popular full-text journals service. There are advantages as well as drawbacks in each case. In the case of electronic journals, either sourced directly from the publisher or through a second party, the users have the advantage of accessing the journals direct. The disadvantage is that in the electronic journal services, search facilities are kept at a low profile. Whereas in the aggregated services, search facilities are fairly high, with low priority for content layout, get up etc.

The IIM Consortium

The year 2000 was a landmark in the history of IIMs, with the formation of the IIM Consortium. The objective was to ensure among the IIMs, optimum utilization and enhancement of the resources, and to minimise the expenditure by consortia based subscriptions to the commonly subscribed databases and journals. The idea was to approach publishers of CD-ROM Databases to begin with, as a consortia, for better pricing and services. Eventually, other digital databases and journals were also planned to be covered by the programme. The consortium meetings were proved to be very productive and successful. During the first meeting itself, a host of databases were jointly purchased at very competitive prices, and a number of others promised supply of their products at a nominal cost.

In the case of journals, all the six IIMs put together subscribes to over 2550 scholarly titles of which around 1200 are duplications (overlapping titles). Among these, 33 titles are being subscribed to by all the IIMs. Having convinced on the dire need for a journals consortia, major publishers such as Elsevier [11], Kluwer [12], John Wiley [13], Blackwell [14], Taylor & Francis [15], and MCB University Press (Emerald Full-Text Intelligent Library) were approached and they all now participates in the Consortium. The end result has been highly praiseworthy, that over 700 E-journals IIMs are able to get online access, across all the IIMs, by paying a nominal additional amount. The table below shows the salient features of the digital subscriptions at IIMK based on the IIM Consortium.

IIM Consortium Resources

With the formation of the IIM Library Consortium since 2000, it has been growing from strength to strength. The IIMs now have a strong digital information base in the light of the content license (electronic access) it has been able to obtain under the IIM Consortium.

The IIM Consortia has the following information base:

1. Two aggregated full-text E-Journal Databases, viz., Business Source Premier of EBSCO (approx. 3000+ journals) and ABI/Inform of PROQUEST (approx. 2000+ journals), having a wide journal coverage of around 4000 unique titles;

2. About half a dozen bibliographic, Company/Industry (corporate) databases (national as well as international), and value added information services;
3. Over 700 E-Journals covering most of the outstanding publishers in Management such as Blackwell, Elsevier, Kluwer, Taylor & Francis etc. are already under subscription and John Wiley, MCB University Press etc. are underway, to take the number of titles to more than 1000 by 2004.

This scholarly resource base is a vast wealth of information and is highly vital, invaluable and voluminous that it is essential for the Consortia to ARCHIVE THE DIGITAL RESOURCES for future use by the IIMs.

Further, the IIMs are now part of the INDEST CONSORTIUM, a "Consortia-based Subscription to Electronic Resources for Technical Education System in India", set up by the Ministry of Human Resource Development (MHRD). The following resources are available to IIMs under this consortium:

1. IEL Online Library (175 journals)
2. ABI / INFORM Complete (2004 journals)
3. ACM Digital Library (30 journals)
4. CRIS INFAC
5. ASIAN CERC / INSIGHT
6. INDIA INFORMER

Thanks to the Ministry HRD and the INDEST National Steering Committee, the information support from the INDEST Consortium is fortunately gaining strength, with the formation of the Special Interest Group - Management Schools.

Need for Archiving the Digital Content

Long-term preservation and archiving of digital information is an unsolved problem for libraries that need continuing access to information over extended /periods of time, for researchers who rely on a cumulative record of data and scholarship, and for libraries, archives, museums and other cultural institutions whose mission is to preserve knowledge. Almost all information generated today originates in digital form, including scientific research data, scholarly publications, records of administration and financial transactions, and even creative expressions in music and art, but few organizations have the means to preserve digital information for the long-term. Despite significant recent advances in the areas of repository architecture and design, standards, and metadata, there is apprehension that these valuable digital information will remain available and usable in the future.

At the moment the IIMs get only content access facility for the prescribed subscription period, which is usually done on an yearly basis, and the publishers provide access to the content, only for the period for which subscription is arranged. The present situation is that the publishers keep changing their information access policies in tune with their changing business policies and

technology capabilities. It is also observed that significant numbers of mergers and acquisitions occurred during the recent past and this also adds up concerns. Most of the publishers park their gigantic digital contents with third party server farms and there is no guarantee that the arrangement will last long.

Key Issues

As already mentioned, the archive is envisaged as a robust backup of the vast digital wealth of the IIM Consortium and in addition, a central repository playing the role of an intellectual network nerve centre. The archive should ensure perpetual as well as uninterrupted access to the licensed content by all the IIMs. The key issues are sourcing the contents and the creation of the archive as a long term preservation solution. Sourcing the content whether raw or otherwise along with structured metadata in XML format is very crucial for a meaningful resource discovery. Full-text searching facilities capable of penetrating down to the paragraph level are prerequisite to an archive of this stature. Content migration strategies between different versions and platforms are serious considerations. Arriving at the access architecture is something which come up with ones own experience and hard work. Open source approach is recommended owing to its relative strength in the e-publishing ecosystem as against proprietary standards.

The Approach

The required infrastructure for hosting the Archive service will be set up. A high power archive server shall be commissioned for the purpose, supplemented with Internet connectivity. The server shall have access to a high end Hard Disk Drive of around 2 TB capacity. The Server will have strategic configurations such as dual processing and mirroring capabilities. A suitable Archive Software will be identified to host the data to be sourced from publishers. For information retrieval a user-friendly search interface shall be developed in HTML/ASP/PHP and placed in the Home Page of the Archive. The database shall be accessed via CGI Scripts and other scripting languages with the help of an HTML/ASP/PHP front-end. XML based metadata description support will be provided to enhance the retrieval capability. The archive shall act as the central repository of the IIM Consortium resources. A state-of-art Fire-wall software will be provided for efficient data security.

Data capturing shall be done jointly by the IIM Consortium. The project Coordinator at IIMK shall be in touch with the journals / databases Coordinator(s) of the IIM Consortium for Archive Data, from time to time. Data for archiving, current as well as retrospective, shall also be part of the forthcoming negotiations/subscriptions of the IIM Consortium. All the participating Publishers will be contacted and negotiated accordingly, by the IIM Consortium.

On an experimental basis, Open Source Archive Software like "LOCKSS" (<http://lockss.stanford.edu/>) and "EAS" (Eprint Archive Software - www.eprints.org) or Open Source Digital Library software like Greenstone (www.greenstone.org) and "Dspace" (www.dspace.org) will be tried, which itself shall save substantial amount for the project. The Open Source systems generally lack full-fledged service support and hence sufficient time may be required for its downloading & costumization, configuration & commissioning, familiarization, and operation & management.

References

1. Terry, Ana Arias (2000). *How today's technology affects libraries' collection choices*. Computers in Libraries Vol. 20, No. 6, pp. 51-55.
2. World Wide Web Virtual Library- Electronic Journals : www.edoc.com/ejournal
3. JSTOR : www.jstor.org
4. Ingenta : www.ingenta.com
5. ProjectMUSE : muse.jhu.edu
6. Emerald Library : www.emeraldinsight.com/ft
7. CatchWord : www.catchword.com
8. J-Gate : j-gate.informindia.com
9. EBSCO Publishing : www.epnet.com
10. Proquest Information & Learning : www.umi.com
11. Elsevier Science Direct : www.sciencedirect.com/web-editions
12. Kluwer Academic Publishing : www.kluweronline.com
13. John Wiley & Sons : www.wiley.com
14. Blackwell Scientific Publishing : www.blackwellpublishing.co
15. Taylor & Francis : <http://taylorandfrancis.metapress.com>