

Program

electronic library and information systems

Migration of data from one library management system to another: a case study in India

Ram Kumar Matoria

Library and Information Services Division, National Informatics Centre, Department of Information Technology, Ministry of Communications and Information Technology, Government of India, New Delhi, India

P.K. Upadhyay

Library and Information Services Division, National Informatics Centre, Department of Information Technology, Ministry of Communications and Information Technology, Government of India, New Delhi, India



Program: electronic library and information systems publishes material on all aspects of the management and use of information technology in libraries and archives, museums and galleries, information centres, and the information industry generally.

Program: electronic library and information systems is published in association with Aslib, The Association for Information Management

MANAGING EDITOR

Lucy A. Tedd

Department of Information and Library Studies, University of Wales, Aberystwyth, UK

REVIEWS EDITOR

John Eyre

Formerly Principal Lecturer, School of Information and Communication Studies, University of North London, UK

PUBLISHER

Diane Heath

ISSN 0033-0337

© 2005 Emerald Group Publishing Limited

Program: electronic library and information systems is indexed and abstracted in: **Scopus**



Certificate number: 1985/99

Awarded in recognition of Emerald's production department's adherence to quality systems and processes when preparing scholarly journals for print

This journal is also available online at www.emeraldinsight.com/prog.htm

Internet services available worldwide at www.emeraldinsight.com

Emerald Group Publishing Limited

60/62 Toller Lane, Bradford BD8 9BY, United Kingdom

Tel +44 (0) 1274 777700

Fax +44 (0) 1274 785200

E-mail help@emeraldinsight.com



INVESTOR IN PEOPLE

Regional offices:

For North America

Emerald, 875 Massachusetts Avenue, 7th Floor, Cambridge, MA 02139, USA

Tel Toll free +1 888 622 0075; Fax +1 617 354 6875

E-mail america@emeraldinsight.com

For Japan

Emerald, 3-22-7 Oowada, Ichikawa-shi, Chiba, 272-0025, Japan

Tel +81 47 393 7322; Fax +81 47 393 7323

E-mail japan@emeraldinsight.com

For Asia Pacific

Emerald, 7-2, 7th Floor, Menara KLH, Bandar Puchong Jaya, 47100 Puchong, Selangor, Malaysia

Tel +60 3 8076 6009; Fax +60 3 8076 6007

E-mail asiapacific@emeraldinsight.com

Customer helpdesk:

Tel +44 (0) 1274 785278; Fax +44 (0) 1274 785204;

E-mail support@emeraldinsight.com

Web www.emeraldinsight.com/customercharter

Orders, subscription and missing claims enquiries:

Tel +44 (0) 1274 777700; Fax +44 (0) 1274 785200

Missing issue claims will be fulfilled if claimed within four months of date of despatch. Maximum of one claim per issue.

Reprints and permissions service:

Anne-Marie Thorslund

Tel +44 (0) 1274 785080

E-mail athorslund@emeraldinsight.com

Web www.emeraldinsight.com/reprints

www.emeraldinsight.com/permissions

No part of this journal may be reproduced, stored in a retrieval system, transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without either the prior written permission of the publisher or a licence permitting restricted copying issued in the UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center. No responsibility is accepted for the accuracy of information contained in the text, illustrations or advertisements. The opinions expressed in the articles are not necessarily those of the Editor or the publisher.

Emerald is a trading name of Emerald Group Publishing Limited

Printed by Printha Group Ltd, Scirocco Close, Moulton Park, Northampton NN3 6HE



PROG
39,2

160

Received 11 November 2004
Revised 31 January 2005
Accepted 1 February 2005

Migration of data from one library management system to another: a case study in India

Ram Kumar Matoria and P.K. Upadhyay

*Library and Information Services Division, National Informatics Centre,
Department of Information Technology, Ministry of Communications and
Information Technology, Government of India, New Delhi, India*

Abstract

Purpose – To share the experiences gained during the migration of library data from one library management system (LibSys™) to another (e-Granthalaya™).

Design/methodology/approach – The paper describes the step-by-step approach taken to migrate the existing library data to the new software. The paper also discusses the peculiarities of the source software from which data were converted and the practical approach adopted in solving the problems faced during the conversion.

Findings – During the process of the conversion of the library data from one software to another many lessons have been learned. These lessons and experiences will be useful for us in the future to import/export the data from other software such as MINISIS, CDS/ISIS, TECHLIB PLUS, etc. being used in Indian libraries. Moreover, the switch over from one software to another is also useful for the libraries as the existing data are refined and cleaned during the conversion process.

Originality/value – The authors share the lessons learned during the conversion process and hope that these will be useful to others in migration of data.

Keywords Computer software, Management techniques, India, Data handling

Paper type Case study

1. Introduction

The data in the library management systems typically include the catalogue records, holdings information and details of who has borrowed materials from the stock and these are generated as a result of various information processing activities in libraries. The ability to migrate data is a key aspect of any change from one library management system to another. The existing library data is an important asset and cannot be generated again and again as it is costly in terms of resources such as money, manpower and time. Therefore, it is imperative to re-use the existing data with the new library management software by converting it in a way suitable for the new software. This is the solution adopted by many libraries (Reuben, 2003).

The authors gratefully acknowledge Dr N. Vijayaditya, Director General, NIC who gave approval to publish the case study in an international journal. They also acknowledge the moral support and guidance always received from Sh. M Moni, Deputy Director General, NIC to undertake this and similar projects to automate the Indian libraries with eGranthalaya™ wherever requested from Government libraries. They also thankfully acknowledge the support provided by the ALTTC Officials, especially Sh. Ashish Tayal, DGM(TX-II); Sh. S B Singla, DE(WLL) and; Sh. Vishnu Bhagwan (In Charge, Library).



Converting library data from an existing library management software to another is not a simple and painless process and can provide many challenges. The case study described here outlines the experiences at the Advanced Level Telecommunication Training Centre (ALTTC) in India in migrating data from the Libsys™ library management system to the e-Granthalaya™.

2. Library management systems at ALTTC

ALTTC is a department of Bharat Sanchar Nigam Limited (BSNL) and is situated at Ghaziabad in Uttar Pradesh in India. ALTTC is a premier telecommunications institute which provides manpower training for BSNL. The library and the institute was established in 1975 (see <http://alttc.nic.in/fsabus.htm> or www.alttc.bsnl.co.in/about_us.htm) and houses about 18,000 books, manuals, standards, etc. and subscribes to about 100 technical and general magazines/periodicals.

The library has used the LibSys™ library management software since 2002 and has created about 13,000 bibliographic records in electronic form. LibSys™ is the most popular library management software in India and is currently being used in more than 90 per cent of the libraries which have an automated system. This software has been designed and marketed by M/s Libsys Corporation, New Delhi (www.libsys.co.in). The software has many version/editions for various kinds of libraries, operating systems, and so on. The version being used in the ALTTC Library prior to conversion was version 4.0, released in 2000 and operating on a Windows platform.

LibSys™, being a “proprietary” software, does not provide access to source code for users and thus data can only be manipulated with a front end program. The version used by the ALTTC library does not use a relational database management system (RDBMS) at its back end, rather, it uses a “flat” file structure where various data elements are stored in the machine readable catalogue (MARC) format using MARC tags and field/record separators. Although LibSys™ has an inbuilt import/export interface, it is not always possible to acquire the existing data in the required format. Moreover, LibSys™ does not permit the export of acquisition and circulation data and restricts the export of catalogue and holdings data to the following file formats (Ghosh, 2001):

- LibSys format;
- MARC format;
- ISO format; and
- text file (fixed/variable length).

There are a number of reasons why any library may wish to switch over from one library management systems to another. Some of these reasons are listed below:

- phasing out of the existing software by its vendor;
- change of version by the vendor;
- lack of required features in existing version/software;
- cost of maintenance of the existing software;
- change due to the hardware/software environment/platform; and
- change forced due to the technological advancements.

In the case of the ALTTC Library, the cost of the annual maintenance contract for the LibSys software was high and new software was being made available, for free, along with training, support and data conversion facilities from the National Informatics Centre (NIC) Department of Informatics, Ministry of Communications and Information Technology, Government of India. This new software, called e-Granthalaya (ver.1.0, release 2003) is a library management system designed and developed by NIC. The current version of e-Granthalaya (eG1) is being distributed free to government libraries under the e-Governance movement of the Government of India. The software operates in Windows environment and uses Microsoft's SQL Server and Microsoft Access at its back end to store the bibliographic details of various reading materials in libraries. The current version does not have any import/export interface, but version 2.0 (eG2), which is under development, will include such utilities.

3. Process of data conversion

3.1. Study of the existing system/data design

The data conversion process started with a study of the existing software operations, the volume of the data available, the data format and the availability of import/export features. In fact, both the authors (who are from NIC) were well versed with the existing version of LibSys™ as they had experience of using it in their library between 1988 to 1998. The authors visited the ALTTC library, viewed the sample of data and tested many other options related to export of the data.

3.2. Export of the existing data

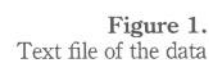
The first step in data migration was to map the existing data into a suitable format from LibSys™. As LibSys™ can generate output files in various formats it was decided to generate variable length text files of the data, which could cover catalogue fields as well as the holdings fields in a single record. We opted for the text output format as it is easier to manipulate data from text file to the eG1 back end database which is the Microsoft SQL Server. Data in a "flat" file structure such as MARC, ISO and other international exchange formats have some limitations when being imported to RDBMS as it is very tricky to maintain multi-relationships among the data elements while importing the data. XML is the latest solution as exchange format but the existing version of LibSys™ does not provide an XML interface. During this process we took the output as a text file covering all the fields in the catalogue and holdings records, such as:

- title;
- authors;
- edition;
- editor;
- corporate author;
- place;
- publisher;
- year;
- pagination;

- ## Migration of data

After generating the data in the text file, we then manipulated/reformatted it using various common utilities, such as the global “Replace” function in Word/Notepad. We then used the # sign as the field separator and the line break as the record separator and trimmed the extra space between the words, if available. An example of the Text file for the data is shown in Figure 1.

We decided first to convert data elements from the Text file to an *ad hoc* database designed in Microsoft Access™ because it was very easy to refine the data using various features of the MS Access. We created tables in MS Access™ that were a



replica of the destination tables and defined various attributes of the fields such as length, data type, and so on. We rechecked data and changed/edited if required, removed unwanted/junk characters, if there were any and checked the spellings of the data elements using sorting/filtering utilities of MS Access™. For extracting relevant data from Text file to MS Access™ we wrote a small utility using Visual Basic™ 6.0.

3.5. Data modification and rationalisation

After conversion of the data from delimited Text to an MS Access™ table, it was necessary to generate various Authority tables such as "Authors" and "Publishers" from the *ad hoc* table. LibSys™ stores "Author" and "Publisher" data with each and every catalogue record repeatedly as it has a flat file structure, while the RDBMS utilises primary keys to link relational data files. Therefore we needed to store such data once in the master tables which were then re-used by storing the pointer/id/primary key with other tables to establish relationships. For this purpose, we again used utility written in Visual Basic which extracted "Distinct/unique" names of "Authors" and "Publishers" and stored them in the respective master tables ("Authors" and "Publishers"), thus, generating the authority tables. In this process the data were refined and various local adjustments were made to rationalise the data.

3.6. Transfer data from ad hoc table to SQL Server tables

After proper testing with sample data, we managed to convert the complete text file to MS Access™ tables which were finally transferred to the MS SQL Server™ database by using the import/export utility of the latter. The database in MS SQL Server™ as back end for eG1 is an integrated database and comprises more than 39 tables and 13 views with multiple relationships among the tables. As we converted only catalogs and holdings data from LibSys™ which were accommodated in the six inter-related tables in eG1 database ("Cats", "Holdings", "Publishers", "Authors", "Subjects" and "Library"). As a matter of fact, circulation data was not recovered as LibSys™ does not provide import/export facility for it.

3.7. Testing of the data

After the data was transferred to the SQL Server™ database which was being used as back end for the eG1 we tested the functions/utilities of the software and then finally implemented the full system in the ALTTC Library. There we tested the system again and handed over the system to the ALTTC authority. Now, the new library management system in the ALTTC library is being used for circulation and other functions successfully. eG1 provides a web interface for users to access and search the library catalogues using any internet browser.

4. Cost of the conversion

The cost of the conversion, in terms of man hours spent for writing the data conversion program in Visual Basic™ and carrying out the full conversion process and other resources used at NIC, was 1.8 lakh rupees (or 180,000 rupees or about £2,197). This cost does not include the cost of the eG1 which is distributed free to Government libraries in India. We managed to convert all the 13,000 records successfully using the methods outlined above and saved a substantial amount of money compared to the cost of re-keying all the data for the new software.

5. Lessons learned

During the conversion process of the library data in the ALTTC Library we observed the following points:

- In general, libraries should avoid purchasing software with a “proprietary” database which results in depending on the software vendor for data migration.
- Data conversion is very easy from/to RDBMS where data can be manipulated directly without the need for a front end program.
- Various international exchange formats such as MARC, Common Communications Format, ISO, etc. are not fully compatible with an RDBMS database as the former have a “flat” file structure making relationships with other data elements difficult.
- The library community should make use of the XML as exchange format due its compatibility with RDBMS as well as the Web environment where multi-relationships can be maintained easily.
- While converting the data from one system to another, sometimes we need to use an intermediary program for manipulation of the data in parts or full. For such a program, Visual Basic™ is the best option for the front end program while Microsoft Access™ is appropriate for the back end solution, as it provides certain utilities to sort, find, replace, copy the data.
- During the data conversion, some of the records may get lost in part or full due to bad output from the source data; absence of the key data fields such as title, etc; availability of junk characters due to non-compatibility of the fonts and so on.

Moreover, converting data from one system to another requires proper planning, use of various IT tools, various local adjustments to the existing data to accommodate the new software. Though, there are a number of ready-made conversion tools available (de Keyser, 2000), these cannot always be used with each and every library management software. These are the issues considered while converting data from LibSys™ to eG1:

- nature of the library management software;
- availability of the import/export formats (MARC, CCF, ISO, TEXT, HTML, XML, etc.);
- data elements available (fixed/variable length);
- field/record separator;
- relationships among the data elements; and
- date, currency, etc.

6. Conclusion

During the process of the conversion of the library data from one software to another we have learnt many lessons. These lessons and experiences will be useful for us in the future to import/export the data from various other software such as MINISIS, CDS/ISIS, TECHLIB PLUS, etc. being used in Indian libraries. Moreover, the switch over from one software to another is also useful for the libraries as the existing data are

refined and cleaned during the conversion process. This certainly enhances the system efficiency for both library users and library staff.

References

- Ghosh, S. (2001), "Data import and export through LibSys: a case study", *SRELS Journal of Information Management*, Vol. 38 No. 4, pp. 309-16.
- Keyser, P. (2000), "Conversion to and from CDS/ISIS: an overview of the possibilities", *Program*, Vol. 34 No. 2, pp. 159-76.
- Reuben, E. (2003), "Migrating records from proprietary software to RTF, HTML and, XML", *Computers in Libraries*, Vol. 23 No. 6, pp. 30-4.

Further reading

- Feather, J. (1991), "The preservation of information: principles and practice of format conversion", *Library Review*, Vol. 40 No. 6, pp. 7-12.
- Sweeney, R. (1996), "Retrospective conversion of library catalogues in UK institutions of higher education: a quantitative analysis", *Library Review*, Vol. 45 No. 2, pp. 52-7.
- Voorbij, H. (2001), "A feasibility study of the retrospective conversion of the Dutch Union Catalogue", *Program*, Vol. 35 No. 1, pp. 57-66.