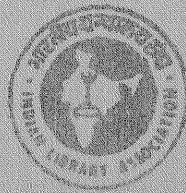


LIBRARIES & INFORMATION SERVICES
IN THE
ELECTRONIC INFORMATION ERA



ILA SEMINAR PAPERS
44TH ALL INDIA LIBRARY CONFERENCE
HYDERABAD, 25-28 FEBRUARY 1999

EDITOR
PROF J L SARDANA

ASSOCIATE EDITORS
M L SAINI S M DHAWAN
S MAJUMDAR DEV RAJ SINGH

INDIAN LIBRARY ASSOCIATION

ROLE AND IMPORTANCE OF DATABASE CREATION IN THE TRANSITION TO EMERGING ELECTRONIC INFORMATION ERA: A CASE STUDY

D K Tewari*
Rajpal Walke*
S M Dhawan*

ABSTRACT

This paper gives an account of retroconversion activities undertaken in the NPL Library. It describes the approaches undertaken for the purpose and discusses their usefulness. This case study highlights the role of library databases in the transition of traditional library to electronic era.

1. INTRODUCTION

Now a days, libraries world over are increasingly using computers and related information technologies to improve efficiency and effectiveness in providing their information services. These technologies have come to catalyse information scenario and have led to development of library networking for resource sharing. Libraries are seeking cooperation at local, regional, national, and international levels for implementing resource sharing programmes. Databases form the basis for sharing library resources in the distributed information environment. In this context, creation of databases, particularly of cataloguing records, is perceived as a pre-requisite in the transition to electronic era. At the National Physical Laboratory Library, an attempt has been made for creating a database of cataloguing records of its monographs. This paper gives a detailed account of retro-conversion activities undertaken for the purpose. It also debates various approaches undertaken for the purpose and discusses their usefulness.

2 NPL LIBRARY

The National Physical Laboratory (NPL) is the earliest the chain of laboratories set up under the Council of Scientific and Industrial Research (CSIR). It was established in 1950 with the objective of strengthening and

* NPL Library, National Physical Laboratory, Dr K. S Krishnan Marg, New Delhi 110012

advancing physics based research as well as physics based technologies in the country. It has since emerged as a leading research organization in the country.

In organisational infrastructure of the Laboratory, NPL Library constitutes as its important component. The Library was created in 1950 with the express purpose of providing information support to scientific community in the Laboratory for research and development work. It has since developed a rich collection in physical sciences with particular emphasis on physics literature. Monographs and periodicals constitute the bulk of its collection.

3 METHODOLOGY

3.1 Software Selection

CDS/ISIS has been used for database creation. It is a versatile software. It comprises a set of programs which offer functionality and performance in database creation. The specific reasons which influenced its choice for the purpose are stated as follows:

- 1 CDS/ISIS allows for efficient use of the memory space on disc since it provides for variable field length fields in database creation. Comparatively, other database management systems do not offer such a feature.
- 2 It offers powerful search features for precision and recall.
- 3 It is able to create and manipulate an unlimited number of databases simultaneously.
- 4 It offers integrated programming facility using Pascal language. It helps in developing specialized applications.

3.2 Data Capturing

Retrospective conversion of library holdings in machine readable form started in 1996. NPL Library did not have the sufficient manpower to carry out this project in-house. It was, therefore, decided to engage an external agency for the purpose. For reasons of performance and quality, we decided that the external agency would be engaged for inputting data in respect of descriptive portion of cataloguing records only. The remaining portion of cataloguing record pertains to subject headings. Since it requires certain amount of professional skill and expertise to verify subject heading data for accuracy and comprehensiveness, it was decided to input this data in-house only.

For capturing data for the purpose of retro-conversion, three options were considered. These include:

- 1 Capturing data straight from the document
- 2 Capturing data from accession register
- 3 Capturing data from catalogue cards

Capturing data straight from the document is indeed a very useful option. It permits capturing data for the purpose as accurately and comprehensively

as possible. However, such an approach was not considered feasible since it would amount to withdrawing documents for circulation, thereby leading to disruption in the library service to users. Capturing data from the accession register was also not found feasible since it does not offer complete data for the purpose. Data such as class number, complete details of authors, and series title are usually missing or incomplete in the accession register record. Given these circumstances, it was decided to adopt the last option for data capturing. The cataloguing data was, therefore, captured from the cataloguing record. Wherever it was found that the data was incomplete or inaccurate, it was supplemented by going to the document.

Data capturing was undertaken by subject. The priority was given to the most used publications in the library. Therefore, publications under class 53 were taken up first for data entry, followed by publications under class 54, 61, 62, and so on. Once publications under all classes are over, the left over publications, if any, would be identified by using accession number key and data undertaken for completeness of the database.

3.3 Hardware Used

The NPL Library is equipped with five personal computers. For the purpose of database creation, two computers were used: PC486 system and PC386 system.

3.4 Standards Used

Common Communication Format (CCF) provides a common method for structuring a bibliographic record so that records could be freely exchanged among different information systems. Data entry worksheet has been designed following the CCF tags. The bibliographic contents in the record were described by using the Anglo American Cataloging Rules(AACR-2). The subject headings were created by using *Thesaurus of scientific technical and Engineering Terms*, 1988 and the abridged edition of UDC.

4 DATABASE STRUCTURE AND DESIGN

The database structure has been worked out by defining the following worksheets and tables for the field definition table (FDT), data entry worksheet, display format, field select table (FST)

4.1 Field Definition Table

The FDT basically defines the fields and their characteristics like TAG - a unique numeric value identifying the numeric value, field name, field length, field type i.e. alphanumeric, numeric, alphabetic pattern etc.- repeatability and subfields. The fields created for our database were in conformity with the CCF. The FDT has been described in Table 1.

Table 1: Field Definition Table for Library Database

TAG	NAME	LENGTH	TYPE	REP	DELIMITER/PATTERN
1	REC.I.D.	15	X		
15	BIB.LEV	5	X		
20	SOURCE OF REC	25	X		ABCDL
21	COMPLETENESS	5	X		A
22	DATA ENTERED ON	15	X		A
30	CHARACTER SET	5	X		ABCDEFG
31	LANG/SCRIPT	20	X		AB
40	PHYSICAL MEDIUM	20	X		AB
80	LINK TO MONOGRA	10	X		AB
100	ISBN	50	X		ABC
200	TITLE	300	X		ABLS
210	PTITLE	200	X		ABLS
240	UNIFORM TITLE	150	X		ABLS
260	EDN.RES.	80	X		ABL
300	100	100	X		ABCDEFZ
310	CORPORATE AUTH	150	X		ABCDEFGLSZ
320	NAME OF MEET	150	X		ABCEGHIJLSZ
330	AUTHOR AFFILIA	100	X		ABCDEL
400	PLACE & PUB.	150	X		ABCD
440	DATE	20	X		AB
460	COLLATION	100	X		ABCD
465	PRICE & BINDING	25	X		ABC
480	SERIES STATEMENT	125	X		ABCDLS
490	PART STATEMENT	50	X		ABC
500	NOTE	200	X		A
510	NOTE BIB.LEVEL	50	X		A
530	CONTENT NOTE	500	X		A
600	ABSTRACT	55	X		A
610	CALL NO.	30	X		AL
620	KEYWORDS	400	X	R	AB
900	HOLDING DETAIL	125	X	R	AB
980	SUBJECT	35	X	R	A
990	COLLABORATORS	50	X	A	
991	ENTERED BY	2	X	A	
992	SORT CODE	10	X		

4.2 Data Entry Worksheet

An input data entry worksheet has been created. The data entry was done by our approved vendor. First, the vendor was asked to fill data under various fields. After verification of the filled in worksheets the data were entered into the machine by using ISISENT menu. The descriptors were entered later in-house by using *Thesaurus of Scientific, Technical, & Engineering Terms*. The worksheet included the necessary fields for data inputting. The keys for sorting and searching were also defined in designing the worksheet. Complete details of the data entry worksheet have been given in Table 2.

Table 2: Data Entry Worksheet

REC.ID. _____ BIB.LEVE _____

SOURCE OF RECORD _____ DATA ENTERED ON _____
 COMPLETENESS _____

LANGUAGE/SCRIPT OF RECORD _____ CHARACTER
 SET _____ LINK TO MONOGRAPH _____

PHYSICAL MEDIUM _____

ISBN _____ YEAR _____

TITLES/RES _____

UNIFORM TITLE _____

EDN. RES. _____

AUTHOR _____

CORPORATE AUTHOR _____

NAME OF MEETING _____

AUTHOR AFFILIATION _____

PLACE AND PUBLISHER _____

DATE _____ COLLATION _____

PRICE AND BINDING _____

_____ SERIES STATEMENT _____

PART STATEMENT _____

NOTE _____

NOTE ON BIB.LEVEL _____

ABSTRACT _____

CALL NO. _____

KEYWORDS _____

HOLDING DETAILS _____

SUBJECT _____

COLLABORATOR(S) _____

ENTERED BY _____ SORT CODE _____

4.3 Subject Descriptors & Keywords

Descriptors were selected from *Thesaurus of Scientific, Technical and Engineering Terms* published by Science Information Resource Center, 1988. For adding keywords we used the abridged 3rd revised English edition of the UDC. Both the subject and keywords are searchable fields.

4.4 Display Format

Display format has also been defined by using the print formatting language of CDS/ISIS. It consists of field selection parameter, horizontal and vertical

parameters, constant and display mode parameters.

With the help of Display format the records in the database are displayed on the screen. Display format determines the pattern of data extraction. The display format is a string of specification in the form of a formula using letters, numbers and symbols. Each fields is indicted in the format by specifying the field tag preceded by the letter V indicating the variable field.

4.5 Field Select Table

Field select table has been created. The fields which have been made searchable are as shown in Table 3. Inverted files have been created on these fields. CDS/ISIS allows for using five indexing techniques for inverted file. The technique number (1) and (4) were used for searching data at the field level, and technique number (0) for searching data at the subfield level.

Table 4: Data Extraction Format

ID	IT	Data extraction format
6	4	V200^B
2	0	V200^A
3	0	V300^A
5	0	V610^A
1	1	V620+ %
7	1	V900+ %
8	0	V300^B

4.6 Printing Database

Records from the database can be printed by using ISISPRT programme. An example of the output of a database record is given in Table 4.

Table 4: Database Output in ASCII Format

0001

TWO LECTURERS; 1. THE PRESENT SITUATION IN THEORY OF ELEMENTARY PARTICLES ELECTRON THEORY OF SUPERCONDUCTIVITY / .- CAMBRIDGE: CAMBRIDGE UNIVERSITY PRESS, 1949.

51P.

537.312.62

87263

SUPERCONDUCTIVITY.ELEMENTARY PARTICLES.ELECTRON
THEORY.

5 CONCLUSION

Resource sharing succeeds on the availability of cataloguing databases for document identification and location. Obviously, the role of databases is indeed very important in the transition to electronic era. Database creation activity should, therefore, be encouraged at all levels of professional activity.

CDS/ISIS has been found to be a useful software in creating cataloguing database at a low cost. Whenever a library is running short of hands, employing external agency for handling retro-conversion work in part has been found to be a useful strategy. The descriptive portion of the cataloguing records may be handled by the external agency. The subject heading portion of the record may be handled by in-house staff. Such a strategy ensures both quality and performance in database creation, and in information retrieval by subject.

In libraries, CDS/ISIS software is currently employed both for database creation as well as for information retrieval. Whenever a CDS database is opened to public for searching and retrieval, it runs the risk of accidental modifications at hands of unsuspecting users, or even from someone who is well versed in handling CDS/ISIS software system. Taking adequate safeguards in advance is the key to protect the library database. Therefore, the database should be opened to public for access and retrieval only after its file attributes have been changed to read only.