

PROBLEMS OF LARGE LIBRARY DATA BASE CREATION

Dr S R Ganpule

ABSTRACT

It need not be overemphasized that for successful computerisation of libraries, it is necessary to have a sound database. Standardization of records is required not only for efficient retrieval within the library, but also to facilitate communication among the libraries far and wide. However creating such a database is a formidable task both in terms of workload involved as well as in terms of the expertise that is needed. In the days when instant results is what is expected out of a project like library computerisation, the fact that database creation requires time and money - is a proposition very difficult to be digested by all the concerned people. The efforts to show the quick results further complicate already difficult problems. The first and major casualty is the quality of the database. It remains incomplete in contents as well as in standards. Several other complications are seen. This paper proposes to highlight the problems relating to database creation in a library which is unusually big. It attempts to suggest some alternative solutions.

COMPUTERISATION WAVE

Right from the days the computers were made popular in the country, computerisation of libraries has been one of the attractive ideas for the library managements. The discussion about computerisation of library and the aspects related to it has been going on for the last two decades. Although these discussions remained only on the theoretical plane, they made some expertise available at the time when there were signs of library computerisation becoming a reality. Those who were exposed to library services in the advanced countries had the practical experience of using computers in the libraries. Some of the bigger libraries were participants in the information programs at international level. As such they got the advantage of "Transfer of Technology". It was with great ease that computers were introduced in these libraries. This helped in a positive way to establish that the computerisation of libraries can be reality. The decision of the UNESCO to make the CDS/ISIS software available free of charge further facilitated the process. The software talent in which our country has been rich started working for developing library operations softwares and making them available on a commercial basis. The proposal for creating a giant network of universities and college libraries coming from no less important person than the Chairman of the University Grants Commission, really sold the idea of library computerisation in far corners of the country, right upto the taluka level, if not village, where a college existed. The initiatives taken by NISSAT for creating city networks and sectorial information

centres for dissemination of information using computerised technology and showing successful results through them, helped in a long way to create a confidence among all for successful use of computers in libraries. The external environment thus created a powerful force in favour of library computerisation.

THE CHALLENGE

However the major stumbling block was creating database of the holdings of libraries. The libraries being a familiar phenomenon to co-exist with the educational and research institutions for more than a century, there are hundreds of libraries which are very rich in collections. This has made the task therefore a gigantic one.

In the initial stages the difficulties could not be realised as the details that had to go into each record were oversimplified. It was believed that author, title, publisher and date of publication and a few subject headings which are the items familiar to even non-librarians would suffice to have a record. The fact that handwritten or typed catalogues of libraries were available further simplified the issue which boiled down to the level of copying the catalogue entries on the hard disc of the computer and develop a program for retrieval by the fields. An impression, therefore, was created that a data entry operator with the minimum knowledge of basic English and keyboard is what is required. With the advent of scanners it was further felt that even the data entry

operator is also not required as record on each of the catalogue cards could be transformed to the hard disc by one who knows scanning technique.

Soon it was realised that creation of a record of an item of the library material which could be used for information purpose was not a simple matter. There are atleast fourteen decision points for identifying the "author" of the book. For example, there can be more than one author, there can be author known by more than one name or there can be no author at all. So is the case with all the fields. In other words the whole of cataloguing code, be it AACRII or Classified Catalogue Code or any other was required to be used in all its details while creating the record of an item pertaining to the library material. Again it was realised that the library material does not simply consists of books and periodicals but a variety of thirty-plus types such as standards, patents, etc. each of which requires a different treatment. The card catalogues prepared in the libraries were the simplified versions of the standard forms. To prepare such versions the librarians had to have knowledge of the cataloguing rules. The work undertaken by ignoring these facts had to go waste because even though they gave some quick results; they were found to be too inadequate to be used as library database in the real sense of the term.

Although this was realised, the fact remains that the creation of the library database is a difficult task. In order that the record satisfies the minimum requirements, it is necessary that each document is examined by library trained person. All the features, including special ones, of the document have to be picked up, entered in a desired manner, checked by another knowledgeable person. The critical fields are the subject headings. Identifying and describing them in the standard vocabulary requires expertise. For all other fields sufficient knowledge of, if not mastery over AACRII is required. In addition to this, sufficient familiarity with the information seeking pattern of readers of all kinds is necessary. The data-sheet thus prepared can be entered by any data entry operator but they need to be revalidated by another data entry operator if not library professional. This process needs manpower strength of one senior library professional and a data entry operator. The time required for creation of one record in this manner is estimated to be 20 minutes for two above mentioned persons. The work of the data entry operator is faster as compared to one who prepares data-sheet. Experience indicates the ratio of workload of two is 3:1. In terms of cost including stationary and machine time it will be Rs. 10/= at today's rate. Thus for a collection of one lakh the work will involve 33,000 man hours and a million rupees. It will be readily agreed that it is a difficult job to make the authorities willingly accept such a

patience and sanction such a huge amount of money.

ALTERNATIVES

Several activities have been thought of to economize on time and expenditure. The one with which there has been a near-agreement is downloading records from the existing databases. There is almost a consensus on the view that the databases of Library of Congress (or OCLC) and the British Museum would take care of 90% of the holdings of any library in India. The second alternative is to create a database of one large library in India and use it for other libraries. The third alternative is to assign this work to external agency. And the last one is to do it yourself. On examining the pros and cons of these alternatives perhaps it may be possible to come out with a workable solution.

ANALYSIS

Although the first solution of downloading from the databases of large libraries appears to be attractive, it may not prove to be meritorious in the long run. In terms of time the saving would not be more than 50% of time required for other alternatives. For checking record of each book in the library, modifying it to provide for the local call number and subject headings and downloading from the library database will require ten minutes. There will be saving of service of data entry operator but that will be offset by the time required for creating the records of unmatched 10% items. In terms of cost, the prices of CD-ROM databases, the hardware, including drive and CD-ROM tower will be a substantial investment.

The major disadvantage of the second alternative is that the other library will have to wait till the database of feeder library is completed. Even after that the creation of records of unique items in individual library will take some more time. The problems with the contractual work are: there is no cost saving as the contractors profit will have to be added to the basic costs; and there will be less control over the quality. Monitoring of the work will also be difficult.

It will also not be desirable to let each individual library to create its own database because, although it will serve the local problem, maintaining the standard required for exchange of information among the libraries will be difficult.

SUGGESTED SOLUTION

In view of the formidable problems involved in a work, the only way that will work is the co-operative way.

At least five libraries of a national level may participate in this venture. Each one of them will create the databases of their own collection in one or two subject fields. This data will be made available on-line on a network like NICNET or I-NET. The other libraries in the country will continuously download the data as and when it becomes available. Other libraries, of course, will wait till a sizable database creation (say 50%) has been completed in the feeder library. After that other libraries will start checking their own collection with the feeder library. By the time they finish first round, more than half collection can be converted. they will separate the entries in

their card catalogue of those books of which the record have been converted. This will give an indication of the unfinished task. By the time the first round is over the database of the feeder library will get augmented; and this will go on for three rounds. After the feeder library completes its own conversion, other libraries will have indication about the left-out items which can either be done locally or, again, in a cooperative manner.

The task is surely difficult, but it can only be accomplished by hard work, dedication, spirit of co-operation and the national pride.