SEMINARY RANGKAIAN
BIBLIOGRAFI KEBANGSAAN

5 - 7 September 1991

The British Council
Jalan Bukit Aman, Kuala Lumpur

anjuran bersama:
Perpustakaan Negara Malaysia
The British Council
STANDING ON THE SHOULDERS . . .
COOPERATIVE ONLINE CATALOGUING SYSTEMS, BASED ON ESTABLISHED INFORMATION RETRIEVAL SYSTEMS

Bruce Royan, SCOLCAP, National Library of Scotland, U.K.

Abstract: argues that local or regional online cataloguing systems now being developed should exploit MARC data via nationally available information retrieval services rather than developing their own potential requirements files. Considers the problems of this; requirements for local processing (acquisitions and circulation), holding of local data, synchrony and authority control. Concludes that the logical structure to handle such problems maps well with the classical distributed data base approach. The paper is illustrated with references to the relationship between SCOLCAP and BLAISE.

1 INTRODUCTION

It is not many years since the speaker on catalogue automation would feel it necessary to justify the use of MARC (Machine Readable Catalog) records as opposed to inputting records from scratch: the figure of speech commonly employed was that it was a choice between generating your own electricity, or plugging in to the National Grid (Ref. 1). Nowadays, when the use of centrally produced MARC records is more commonplace, connection to this metaphorical network has remained less than direct. MARC-based single libraries in the UK have in general contented themselves with taking what they can of their current cataloguing from the weekly UK MARC tapes, while cooperatives like BLCMP and LASER have built up their own massive Potential Requirements Files on disk (Ref. 2). Meanwhile, the combined total of Library of Congress and UK MARC records is nearing 2 million and is growing by some 200,000 pa. Even bearing in mind the continual decline in the cost of direct-access storage, it is unlikely that the maintenance of an online database of these proportions for each automated cataloguing system could be justified, yet although the chances of any particular record on these files being used by a system are slim, there is no foolproof way of excluding records as potential requirements. In the UK we are fortunate that the British Library have for some 3 years now been offering this entire database online for information retrieval purposes, and it appears this is becoming the pattern in other countries. If the designers of library housekeeping systems relied on online access to this national database as their potential requirements file, more of their energies could be devoted to developing services tailored to the users’ needs. It has long been the policy of the Scottish Libraries Cooperative Automation Project (SCOLCAP) to build its system upon a foundation of the services
provided by the British Library, and these are the shoulders mentioned in this paper's title; Newton said 'If I have seen further, it is by standing on the shoulders of giants' (Ref. 4).

2 LOCAL PROCESSING

SCOLCAP is a cooperative consisting of the National Library of Scotland plus currently some 20 Public, Academic and Special libraries in Scotland and the North of England, whose objectives might be summarized as follows:

- Increased cataloguer productivity via the sharing of the intellectual effort of book description.
- Improved management control of the book acquisition progress.
- Assistance with conversion to the revised Anglo American Cataloguing Rules (AACR 2), scheduled for 1981.
- Improved utilization of the bookstock via faster shelving (an elimination of processing backlogs), more access points and copies of catalogues, and greater facilities for interlending and cooperative acquisitions.
- Provision of a responsive service to an increasing number of libraries.
- Integration with the developing national network.

Some of these objectives may be achieved by the direct exploitation of BLAISE (British Library Automated Information Service) services. The batch mode production of COM catalogues for example is as well done in Harlow as it could be in Scotland, and SCOLCAP has strongly committed itself to using the BL Local Catalogue (LOCAS) service, investing much time and effort in specifying enhancements and improvements in LOCAS software that have made it into a very powerful and flexible tool indeed. Similarly the online storage and retrieval of the catalogue data is far better done centrally; this is the burden of this paper.

But there are other areas where the case for centralized processing is much weaker. Such experience over the last 3 years has proved the inappropriateness of cobbling data entry and editing software together with what is basically an Information Retrieval package, and recent BL developments for the Eighteenth Century Short Title Catalogue (ESTC) project distribute these functions away from the mainframe (Ref. 5).

For SCOLCAP, these functions will be performed on a large minicomputer or small mainframe (proposals are currently under evaluation), communicating with BLAISE via a high speed line.

The same equipment will be used by SCOLCAP to cater for its book ordering/acquisitions needs. This facility is marked by a high transaction rate, considerable library-specific tailoring, and the need for rapid turn-round of hard-copy products; not features to be associated with nationally centralized systems. At the same time the usefulness of sharing descriptive data on new material and management information on suppliers' performance suggests that regional cooperation would be most appropriate.

This is not the case with circulation control, where we believe the much higher transaction rate and the much lower 'shareability' of data points to local processing. As each book is accessioned, SCOLCAP will provide bibliographic data for the appropriate library's local circulation system.

It may eventually be possible for the SCOLCAP team to recommend a
particular local system for this, and to provide online support, but to begin with the data will be supplied on a weekly magnetic tape, in a standard stripped UK MARC format.

3 LOCAL DATA
As suggested above, much of the data a library requires on a book, relating to its acquisition, shelving and circulation for example, is of minimal interest to any other library. It also happens to be the data most subject to change. For these reasons, BLAISE does not hold local data online. Unfortunately, and precisely because the data is volatile, online access to local data is most important to any automated library. SCOLCAP, therefore, will hold online in Edinburgh, the local data for each of its libraries' holdings. The appropriate general data may also be available on the Edinburgh machine (see below) or may need to be retrieved from BLAISE, but in either case, access will be through the local data record (Fig 1):

ACCESS
BY

LOCAL
DATA
RECORD

GENERAL
DATA
ON
SCOLCAP

GENERAL
DATA
ON
BLAISE

Fig 1 - Access to Local and General Data in SCOLCIP.
This record will contain all key information (access in SCOLCAP will be provided by Order Number, Control Number, Author and Title), plus all local data, plus the means of retrieving the general data (either a pointer to the record, or an indication that the data is on BLAISE). If the general data is on BLAISE it is retrieved by the system generating a search on control number (obtained from the Local Data Record). The local and general data records are then combined and a 'virtual record' presented to the user, who need not be aware of where this data has been held.

4 SYNCHRONY
The problems of concurrent amendments leading to different 'states' of the same record was with us in batch processing days and loomed large as online updating became a possibility. It may be thought that with the added dimension of data distribution described above the problem would be exacerbated, but this is not, in fact, the case. Firstly, it must be appreciated that BLAISE provides SCOLCAP with 'read only' access online. Material retrieved from BLAISE for the first time is loaded onto the SCOLCIP (SCOLCAP Cataloguing In Process) file. If derived from LC MARC it may be automatically treated to bring it in line with SCOLCAP practice (Ref 6), and in any case it may be edited as required, but any changes will only take place online on the SCOLCIP file, being reported to BLAISE along with new EMMA (Exta-MARC Material) records via a weekly changes tape. These records are retained on SCOLCIP for the period during which they are most likely to be required online, and are added meanwhile to a special SCOLCAP online union file (referred to as SLP MARC) on BLAISE. This contains all records held on the LOCAS files for SCOLCAP, even though they may be versions of, or even identical to,
records held on other BLAISE MARC files; this unaesthetic redundancy at least reduces the problems of update control in BLAISE. At the end of this period the system checks that each record is available on SLP MARC, then discards the general data from SCOLCIP. It should be noted that since access is at all times via the local data record on SCOLCIP, concurrent access control remains a relatively well understood single-site matter.

5 AUTHORITY CONTROL

Perhaps less easy to cope with, is the related problem of Authority Control. BL have on occasion been criticized for taking the British bibliographic community into AACR2 without a satisfactory automated system to control accuracy and consistency in the use of forms of headings, but it must be borne in mind that until the government cuts of August 1979, such a system (MERLIN) was under active development (Refs 7 & 8). In the absence of such a nationally available system however, SCOLCAP has had to design its own solution.

The SCOLCAP machine will support an Authority File, containing records for every name heading used in the SCOLCAP Union Catalogue. Each record will contain a Name Index Number (NIN), the correct form of the name (with its tags and subfield codes as in a MARC record), associated notes and references and a list of all records linked to this heading.

Cataloguing new books (whether EMMA or obtained from MARC) will involve searching the Authority File for the headings required. If one is found, the NIN is quoted in the appropriate field of the bibliographic record, which the system will automatically expand with data from the authority record. Thus although the authority data is held separately by SCOLCAP, insertions will be passed forward to LOCAS as whole MARC records.

Similarly if a change of heading is required, amendment (under password control) of the authority record triggers the automatic generation of amendment messages for all linked records.

6 A DISTRIBUTED DATABASE

The SCOLCAP system, and its particular relationship to BLAISE, were designed as a response to the requirements of a particular group of libraries. It is gratifying to find that it very neatly illustrates what are seen as the three classical theoretical types of database distribution. These are partitioning, fragmentation and replication (Ref 9).

Partitioning can be defined as the separation of the conceptional database into increments that reside in two or more locations. The SCOLCAP example of this, is the separation of a conceptually integrated cataloguing/acquisitions file into the Cataloguing in Process database (SCOLCIP) and the SCOLCAP Online file on BLAISE (SLP MARC). This partitioning is possible because of a useful feature of library housekeeping data which has been touched on above, and which has been described elsewhere as ‘volatility decay’ (Ref 10). The general data records most likely to be required for consultation or amendment are those that have recently been created or amended, and once through the initial active stage are unlikely ever to be touched again. So all general data is kept online on SCOLCIP during, and for as long as possible after, the acquisitions and cataloguing process, before the discarding mechanism comes into operation.
Fragmentation is the distribution of the data elements of a logical record across more than one location, and the SCOLCAP example of this is the treatment of general and local data. When, as described in the paragraph above, the general data record has achieved a relatively steady state, it is discarded from SCOLCIP and is from then on only available from BLAISE. Although the user continues to think of one catalogue record, it is in fact distributed between the two physical database systems.

Replication can be defined as maintaining copies of all or part of the conceptual database, to reduce telecommunications line traffic and to increase availability and resilience. Studies of the intake of the six original SCOLCAP participants carried out in 1975 yielded data which could be expressed in a manner which will be familiar to students of library science:

![Figure 2: Chances of Satisfaction for different PRF sizes.](https://example.com/figure2.png)

Although the figures themselves may be open to question and will be carefully checked in practice, the implication is clearly that a relatively small portion of the total MARC database will satisfy a relatively large portion of the requirements for cataloguing current intake. Replication of part of the PRF should dramatically reduce the transaction load on BLAISE. One of the research aims of SCOLCAP is to establish an optimum PRF subset, but a first estimate would be that it would include between 6 and 12 weeks of the latest UK MARC, loaded from magnetic tape.

7 CONCLUSION

It may appear, that this paper started out by denouncing the holding of private PRFs and ended by advocating it. This is not the case; the replication of part of the BLAISE PRF is a mechanism to improve total system performance only, and should not disguise the fact that in SCOLCAP eyes there is still only one PRF. In the UK there have historically been various pressures, not least the approach to
funding library automation, that have appeared to encourage differentiation and incompatibility. We are a small country with apparently diminishing resources, and the SCOLCAP team are sure that from now on the only way forward for any cooperative is to build on the achievements of others.

8 REFERENCES

1. ROYAN, Bruce Ed
   ‘An Introduction to MARC’ Tape Slide Presentation.

2. SEAL, Alan
   ‘Automated Cataloguing in the UK: a guide to services’.
   BATH: Bath University Library 1980.


4. NEWTON, Sir Isaac
   Letter to Robert Hook, 5 February 1675.

5. NOERR, Peter L. and Michael C WHITE
   ‘Distributed Cataloguing; What to distribute?’ Proc. 3rd
   International Online Information Meeting.
   LONDON: Learned Information 1979.

6. PRINGLE, Reg and Sandy STEVENSON
   ‘LC MARC off the peg?’ ‘Automatic upgrading of LC MARC
   records.’
   VINE 29 July 1979

7. ROYAN, Bruce

8. ROBINSON, Stuart
   ‘Sleeping Beauty : MERLIN, a state of the art report’ PROGRAM
   Vol 14 no 1 Jan 1980.

9. ROYAN, Bruce
   ‘Distributed database in a library context: the SCOLCAP system’.

10. ROYÁN, Bruce
    ‘SCOLCAP - But I thought that was an on-line system?’
    In SJ. TEAGUE Ed COM systems in Libraries: Current British
    Practice’ Guildford MAGB 1978.