

Merlin for the cataloguer

A. Hopkinson

British Library Bibliographic Services Division

MERLIN is the British Library's advanced computer system. A new MARC-type format has been devised to take advantage of recent developments in computer technology. Records are split up into meaningful elements which are stored separately. Catalogue entries consist of a set of pointers pointing to the elements. MERLIN supports shared cataloguing but allows libraries to have their own private views of any element or group of elements. Standardization is easier to achieve as the structure of MERLIN is designed to make use of authority files. Automatic conversion is possible between the MERLIN format and existing formats.

Introduction

MERLIN is the name given to the British Library's advanced computer system: it stands for MachinE Readable Library INformation. This name has been chosen to stress that the system will support all kinds of library information including library housekeeping information, keywords, abstracts, statistical information and indexing systems, compared with MARC which is purely a system for producing catalogue records, standing as it does for MACHine Readable Cataloguing.

MERLIN's first applications will be in areas where present systems cannot cope; first in subject classification and indexing, in PRECIS, where the advanced nature of the work requires the latest computer database technology and then in the area of cataloguing non-book materials where UK MARC, having been devised originally for books, does not have sufficient definition to deal with items which may constitute sets consisting of different types of bibliographic material. The first applications will therefore be the on-line PRECIS file and the Learning Materials Recording Study for which the British Library and the Inner London Education Authority are coming together to produce a prototype catalogue of non-book materials to explore the possibilities of a national materiography produced by cooperative cataloguing as was envisaged in the report *The Scope for automatic data processing in the British Library*.¹

MERLIN: the bibliographic file

Although MERLIN will eventually support the whole of the British Library's computer work it is in the area of bibliographic records that other libraries will be participating; it is with this area that we are concerned here. The British Library already provides a number of services to other libraries to assist them in producing their catalogues. Libraries may purchase the records produced in the MARC (Machine Readable Cataloguing) format on magnetic tape corresponding to all the entries in the British National Bibliography; these are kept up to date by weekly tapes corresponding to the *Weekly List*² to which libraries can subscribe. Alternatively on receipt of a list of control numbers

(BNB number or ISBN), a search is made of the tape files and the appropriate records are copied to a tape which is sent to the requesting library. A third service is the catalogue production service; libraries select their records by means of a control number and may also submit their own records in MARC format for foreign books or UK publications before 1950 which are not in MARC. Catalogues are then produced for these libraries in a number of different formats. Of these three services it is this last service which the new system will benefit most of all; those records which a library has to create itself will be stored on the computer in such a way that other libraries can see them. These services will continue under MERLIN and will be enhanced in the light of experience which has been gained in the computer departments of the British Library and its predecessor the British National Bibliography Ltd.

The present state of machine readable cataloguing in the British Library
The catalogue services which the British National Bibliography provided were based on the MARC system developed by the Library of Congress of the United States in 1964 and the British Library still uses this MARC format. Essentially, the MARC system involves coding each element of a catalogue entry with a three-digit code called a tag. A personal author's name as main entry is coded 100, as added entry 700; title is 245, publisher 260 and series statement 490. These codes are extended by adding indicators, numbers which denote if the title is required as an added entry or the number of non-filing characters at the beginning of a title. Each element or field of information is further divided into subfields which are coded in a different way. On this format was based the official international standard for the exchange of bibliographic records on magnetic tape ISO 2709. This specifies very broadly the requirement of the tapes for exchange. The Library of Congress is developing a number of formats to this standard, for books, serials and all other kinds of materials, but so far at the British Library we only take exchange records of books. If the British Library were to take records for other kinds of materials on tape it would be necessary for new formats to be devised as the present UK MARC format has only the fields required for cataloguing books. Thus it cannot hold data such as scale on maps very satisfactorily as there is no code set aside for scale in the MARC format. The British Library has not yet begun to use any format other than that for books, though formats for maps, music and serials have been developed. There are good reasons for this. There is a set of programs which produce catalogues from the present format. If different formats were produced, slight alterations would need to be made to the programs to enable them to deal with other tags. And if a catalogue were being produced containing records of different categories of materials, the programs would have to be switched each time a different record were being processed, a very unhappy situation. This means that MERLIN must have a common format for all types of media and materials. For this reason the idea of the MERLIN format was mooted: so that all records can be stored in one sequence in the computer and so that one output program can cope with any type of record equally well, especially as the new AACR will not be biased towards any particular type of bibliographic material.

Automation—the future

If a computer is introduced into any process it should never become its master. The format of the catalogue entry has developed over 400 years and librarians are not going to throw this overboard to suit a computer. The MARC format, though developed only recently, is well established and satisfies the requirements of most libraries which only require catalogues of books and if they have catalogues of recorded music may wish to keep them separate in any case. One of the requirements of MERLIN is that it will allow input and output by means of the UK MARC format, though it will store records by means of the MERLIN format. The cataloguer using UK MARC will put his record into MERLIN either by typing it in on a visual display unit or by means of punched card or tape. The computer will then take each record and by means of a process known as tag translation will convert the tags to the MERLIN format ready to be stored. This will be done by looking up each subfield within a field in a table and finding the corresponding one in the new format. When the record is called up again from the database it will be converted back into UK MARC automatically, as the user will have specified beforehand the conversions he requires and these will be stored in the computer. However if the record is used for producing a catalogue entry, the entry will be constructed directly from the MERLIN format. The aim is of course to produce catalogues according to AACR and ISBD in as efficient a way as possible. UK MARC will suffice for the AACR requirements of most public libraries; but for a national library catalogue or for an information retrieval system which will hold records from which are to be produced different kinds of catalogues and bibliographies a new internal format has had to be devised.

The MERLIN format

It must be remembered that the MARC format was devised for holding records on magnetic tape. In order to extract a record from magnetic tape and get it copied into a file to produce a catalogue it is usually necessary to run all the way through a large set of tapes. Each MARC record has a control number, either ISBN or BNB, and if this is known it is easier to find a record. If the date of the record's creation is known it becomes easier to find the appropriate tape but when that is found the computer has to run through the tape looking for the desired record. Advances have been made in computer technology since 1964 when the MARC format was devised and now records can be held on disk and the computer can find the appropriate point where a record is held, almost instantaneously. A disk can be pictured in the following way. Imagine a record player with not one needle and turn-table, but a set of needles all on arms above each other, each able to play a different disc. On each disc are a number of different pieces of music. Each head can point anywhere on its particular disc. This is how a computer disk works. The 'gramophone records' go round and it only takes one revolution for the heads to find the appropriate point. How they do find the appropriate place on the disk depends on complicated hardware technology, but it is just as easy for a disk to be searched for a name as for a control number, so long as they are correctly specified. Advances have also been made in telecommunications technology. In the past

records had to be transferred from one organization to another on tape, magnetic tape or paper tape. Now it is possible for libraries to communicate directly with the central computer over GPO lines and the records can then be consulted on a VDU without any physical objects being moved. Faster access has interesting possibilities. Instead of storing each catalogue record in a string of characters as on a MARC tape it is better to store each element separately according to its type. This makes retrieval of any record faster in the long run and can assist in the production of catalogues as well as retrieval of individual records. Most important of all it saves space in the computer, an important factor if you want to store six million catalogue records in one computer! In fact fast retrieval makes it essential to store records in different ways and build them up from basic elements. Each element is indexed by the computer and referred to by its address or location on the computer disk. These addresses appear to be used as shorthand, or codes, for the data. Each element whether a complete record or just a field will be indexed in this way if it will be useful for retrieval. Instead of repeating the data in each record, the computer builds up a record consisting of pointers. These pointers make up what is called a packet, in this instance a work packet; there are also subject and name authority packets. Each work packet contains a pointer to every item of data contained in the record of that book. In addition each library that uses that record has a code attached to the record itself and to each item of data in the record that it uses which is different from the standard record. The standard record is that of the British Library or if none has been entered by the British Library the record input by the user

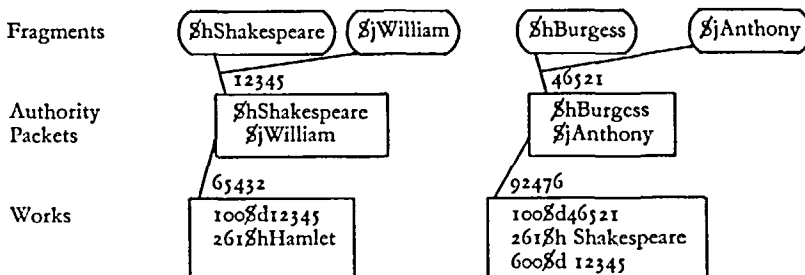


FIG. 1A: Diagrammatic representation of works, authority packets and fragments of text on database

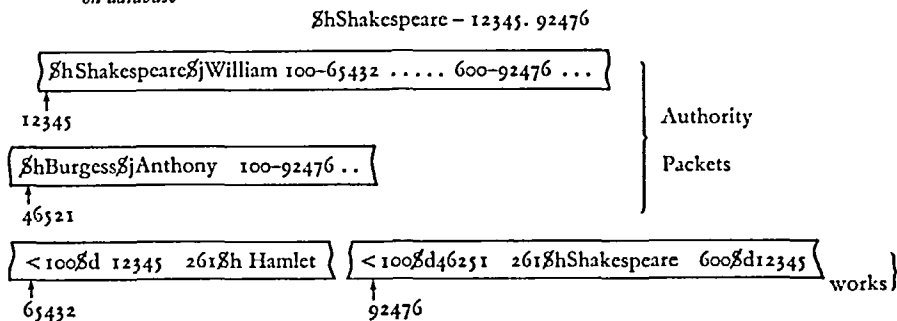


FIG. 1B: Representation of storage of Fig. 1a on database

with the highest authority. Each user will be given an authority level relating to his use of AACR and fullness of cataloguing. So each record has associated with it a list of pointers to the data within it; also each data element equivalent to a field has a list of pointers to records that contain this element and a list of pointers to the subfields within it in certain cases. Thus if a search of the database is initiated for all books by and about William Shakespeare the computer will search its files of data for the name (in MERLIN format $\$hShakespeare\j William) and will then follow up the pointers which will lead to every record containing that name. The role of the name William Shakespeare in the record will be discernible by the tag which will precede the list of pointers and the computer will be able to search selectively depending on the tag. The arrangement of the data is described diagrammatically in Figure 1a. Figure 1b shows how this will be stored in the computer. This diagram does not illustrate that a library can have its own 'view' of any item on the database. This has been left out for the sake of simplicity. If a library wants different information in a record then it will set up another pointer to its own information. It will appear to have overwritten the original information though anyone who requires the original will find it. Unfortunately each 'private view' will take up space and users of the system will have to pay for this; so it will be advantageous for libraries to adopt AACR as it is usually interpreted, or a subset of this.

Large records from smaller ones

One feature of the storage of data in a computer using the most recent database techniques is that larger records can be built up from smaller records. This is done by means of links that indicate that one item is related to another. If a library is interested in a link, then the computer will store that link and mark that link with the code of the interested library. The cataloguer will not need to know about these links; the computer will assume them from the way the cataloguer presents his information. In order to take full advantage of these links it is essential that the data is analysed in a logical way and this is where the MERLIN format differs from other formats. The MERLIN format is not completely devised yet and in fact unlike earlier formats it will be flexible and will always be changing to keep up to date with requirements. The tags are arranged to reflect the order of ISBD. However it is not the order of the tags that makes the important difference but the definitions and these will match those in the new AACR.

Quicker and easier cataloguing

Every cataloguer knows the tedium of cataloguing a large number of books in the same series each of which requires a separate entry. And the more complete the entries the greater the tedium and the amount of time wasted as the cataloguer virtually copies each entry from the previous one making only minor alterations to each one, such as part title and number, and ISBN. The cataloguer at a typewriter might be tempted to resort to carbon paper adding the differences later. The cataloguer at a computer terminal is more fortunate; he can create one record, send it down into the computer and using the same data on the screen can then replace the appropriate words by overwriting them

PHILOLOGICAL SOCIETY

Transactions of the Philological Society.—Oxford : Basil Blackwell.
1972.—1973.—114 p.; 22 cm.

FIG. 2A: *Entry for a part*

PHILOLOGICAL SOCIETY

Transactions of the Philological Society.—Oxford : Basil Blackwell, 854— .
Published in Oxford by Clarendon Press, 1854–1953.

FIG. 2B: *Entry for a serial*

PALMER, FRANK ROBERT

Noun phrase and sentence: a problem in semantics/syntax/by F. R. Palmer.
(In *Philological Society. Transactions of the Philological Society, 1972.*
Oxford, 1973. 22 cm. p. 20–43).

FIG. 2C: *Entry for a serial article*

with new words or spaces. He then has a new entry which he can send down into the computer. In the MERLIN system, items of information are held only once with pointers to them from the individual records. Therefore it will be helpful if the catalogue records can be divided into useful semi-independent units of data, which can be called subrecords. A MARC field with an author's name is one such kind of subrecord, but collections of fields can constitute subrecords, some of which can be brought into a large number of records of books. The text: 'Report on the census of production/ . . . Business Statistics Office—London: H.M.S.O.—(*Business Monitor*)' occurs thousands of times in MARC records. In fact some catalogues would merely require this text once and treat these Business Monitors like a serial, not noting individual parts; in fact one man's record can be another man's subrecord so we call them all records and if a set of these is brought together we call it a record set. These records will belong to different bibliographic levels. In Figure 2a we are interested in the part and so within the record we note the contents and series and concentrate on describing the part. In Figure 2b we are interested in the series as a unit and not so interested in its parts, though in an on-line cataloguing system like MERLIN it will be easy to follow up all the links to the parts. In Figure 2c we are interested in the article and describe this fully, of necessity referring to the serial and part in which it is contained. MERLIN will be able to store the whole record and offer to the user the different parts of the total record that he requires. The record which he wants to see, which will be to him a record in its own right, is called the virtual record. To make a record look complete to different users with different requirements we must analyse the records in a logical way. Whatever type of material or level a record represents, it will have a title and usually author or editor and control number, publisher and date. At any level and within any material each element of description will have the same tag. Each different record in a virtual record will have a code to determine

the record type and each record will have a number between 1 and 99 prefixed by a B to denote bibliographic record (A is an authority record). In Figure 3 each record has a contained/contains relationship with the others, but more especially with the adjacent members of the record set. In addition each record can be categorized, e.g. series, part, analytical, and a code indicating the category will be found in the 001 field. The record number serves to indicate the point of the record in the hierarchy. If there were two articles in the virtual record in Figure 3., they would be labelled B9/1 and B9/2, to indicate they are at the same level. The computer will not store these level numbers but will convert them into links. Each record in Figure 3 would then contain at least one linking field. Records 3, 5 and 7 will contain two, one up and one down. These tags will be in the 700/800 range; 700 will mean contained in and 800 will mean contains. Each tag will be followed by the database number of the work record to which it links. At times it will be necessary for the cataloguer to insert the links himself. If a work is in two distinct series, then it will be clearer if the links are made explicit and it will be essential to make the links explicit if the item in hand has a volume number within one or more of the series, or if one series and not the other is contained within a third and higher series. Figure 4 illustrates this; the cataloguer has made the links only in one direction: the computer will make and store the reciprocals.

Non-hierarchical links

It is only the hierarchic links, denoting contains or contained in, that can be implied. In MERLIN other kinds of links will be available. Because of the way in which MERLIN stores records, advantage can be taken of their links to produce bibliographic history notes. The relationships denoted by MARC tag 503 can be expressed in MERLIN by actually making a link from the work in hand to its related work. Instead of writing out a note at the foot of a catalogue

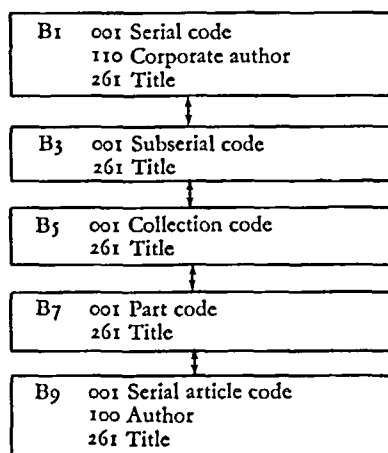


FIG. 3: A record set illustrating an article within a part of a subserial; the part is found in a physical collection of parts bound together (e.g. an annual volume). The subserial is part of a serial

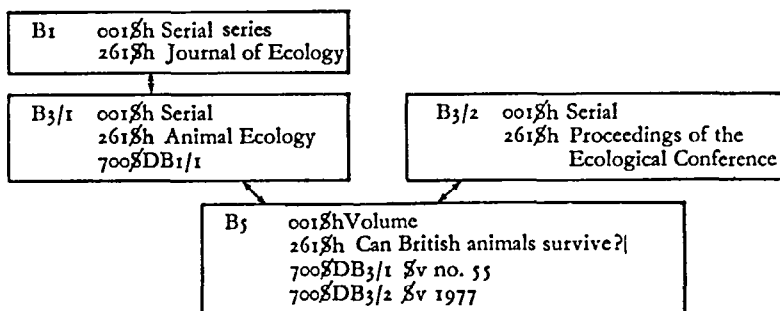


FIG. 4: Links in two hierarchies

entry—e.g. Originally published: 1956—we can make a link to the earlier work when we catalogue the later one. This can be made by quoting an ISBN or other suitable control number in the linking field. Alternatively if the cataloguer is working at a terminal on-line, he may conduct a search for the record he wants to link to and on displaying it on the screen he can give it a record number and add the new record to that record creating a virtual record containing both works. This search parallels the procedure in *British National Bibliography* production at the moment. If a book is found to be a new edition of an earlier work, the cataloguer searches for the entry of the previous work, either through the authority file or earlier volumes of *British National Bibliography*. On finding the entry he uses it to standardize the later entry and takes the date and publisher if it is different from the earlier work. Of course a note like the one above is not required in a catalogue where the works would file together, so this kind of link would not be followed up if the computer were producing a catalogue rather than a weekly list or even annual list of books published. Figure 5 shows an example of a virtual record containing two works, one a new edition of the other. Tag 740 means 'originally published'.

Authority file information

Name authority records will be stored in MERLIN in a similar way to work records. Separate fields will store alternative versions of the name and the references. Other fields will store notes on the authority record. Links will be made between the authority record and the first work which uses the authority

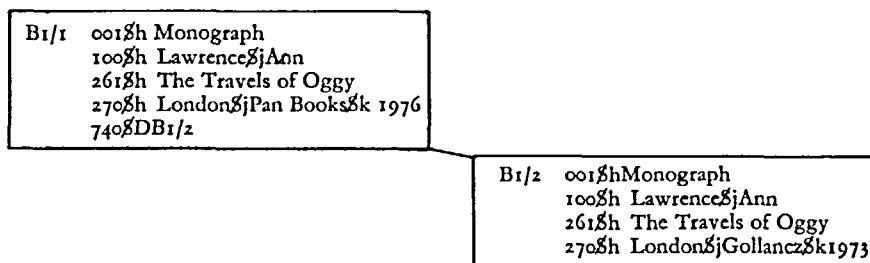


FIG. 5: A work linked to a previous edition of the work

record. These notes and links will serve as aids to the cataloguer to enable him to make a choice between the different authority headings. Where MERLIN will differ from existing systems is that the name that is chosen as the heading in an authority record will be the very same instance of the name used in every record where it appears. This is made possible by the system of pointers. Whenever the name William Shakespeare is used, a pointer will be set to the actual one and only instance of that name on the database. If the name is changed it will be changed in all instances though there will be an overriding mechanism so that if the British Library changes a heading it will not necessarily, if so desired, result in changes in other libraries' records. Furthermore, setting a pointer to a name will result in all the references relating to that name being included in a catalogue that contains that record. Thus it is essential that each name is unique or else if one name were changed the other would be changed automatically. This is no problem as name headings have to be qualified until they become unique. Librarians are fortunate in that the high standard of their records lends themselves to this kind of automation. Once again, cataloguers may change their view of the references by inserting alternative forms in the authority record, which will be marked by their own library's code.

Using the authority file in cataloguing

The cataloguer on entering a work will be able to create a separate authority record for each author in the work record. However, creating a separate record takes time so to avoid this it will be possible for him to enter authors in the normal MARC way, at tag 100 in a record. Added entries for authors will also be at 100 (cf MARC 700) to keep all names together; an indicator will differentiate. The computer will then compare the name as entered with all the name headings in authority records and will set a pointer to the appropriate one. If there exist two headings that match (e.g. if there are two authors of that name which require date of birth to distinguish) the cataloguer will be told either instantaneously on his VDU, or on a diagnostic printout if he is working in batch mode. Ideally the cataloguer working on-line will conduct a search for the appropriate authority records, except in the case of a common well-known heading where there is no chance of a mistake. Having found the name in an authority file search, the cataloguer will be able to insert it into the record by typing in the preferred form as found in the authority record, or by typing in the control number of the authority record (called the name index number or NIN). This will ensure the inclusion of all the necessary references to that heading in the catalogue.

Content of the authority record

The information to be stored in name authority records has not yet been finalized and is awaiting the INTERMARC proposal on authority files. However, so far the requirements of different departments of the British Library have been catered for, and a paper has been written³ and discussed at the British Library Cataloguing Policy Group. There will be the facility for storing in each authority record a note of the AACR rule which has been used in making any decision on the form of the heading, and the source of information for

0XX	Control numbers and coded information
1XX	Names
2XX	Titles
3XX	Media specific information including collation
4XX	Series statement
5XX	Notes
6XX	Subject information
7XX	Links
8XX	
9XX	Private information, e.g. shelf marks

FIG. 6: *Outline of areas in MERLIN format*

creating the heading if outside the work for which it was first created. The nature of MERLIN makes it easy to be able to call up any or all of the works in which a heading is used, without the cataloguer having to do any extra work.

Coding the data within records

Little has been said in this paper about the way the cataloguer will code the information in the records. The tags used will be similar to those in UK MARC. Extra fields have been added; many of these are required for non-book materials and have been found in the LC MARC formats for such materials. Others have been found in INTERMARC, UNIMARC and UNISIST documents and others from the requirements of the British Library and its customers. Because the format is devised with ISBD in mind the subfields do not always correspond exactly with those of the present MARC format. John Linford's paper for EUDISED⁴ was taken as a model for the arrangement of the tags and the requirements of ISBD and the new edition of AACR have been largely superimposed on this, resulting in ten areas as illustrated in Figure 6.

Tag translation

Though these tags are different from the ones used at the moment, cataloguers will not need to change to the new format. Because of a powerful table-driven program in the computer it will be possible to convert automatically existing formats to the MERLIN format, in fact the MERLIN format was designed with this in view. Each user will have stored in the machine a set of profiles which tells the computer amongst other things what kind of output that user requires and what format he uses for input and the validation he requires, so the cataloguer will not need to know of this conversion. MERLIN is already preparing for international exchange of records in different formats with its flexible built-in tag translation program.

Design of MERLIN

MERLIN has been designed by computer people and librarians working together. It has been designed as a completely new system and yet it will be able to interface with existing systems.

A committee of cataloguers drawn from interested parties in the British Library, including the Bibliographic Standards Office has drawn up the format with staff from the computer team suggesting ways in which the format and

the records can be handled. They have kept in close touch with the British Library Cataloguing Policy Group and ideas have been exchanged between the MERLIN team and cataloguers from other departments of the British Library. The foundations of the MERLIN system were laid on the personal experience of members of the team, the experience gained from the Library Software Package and advice from Michael Gorman and the Bibliographic Standards Office.

On the technological side, members of the team have given papers at conferences on information retrieval⁵ and computer database technology,⁶ and written an article in *Program*.⁷

However as a matter of policy the requirements formulated by the cataloguers have overridden ideas of the computer staff where there has been any conflict.

REFERENCES

- 1 LINE, M. B. et al. *The scope for automatic data processing in the British Library*. Part 1, p. 79. HMSO, 1972.
- 2 BRITISH LIBRARY. *British national bibliography: weekly list*.
- 3 BRITISH LIBRARY. BIBLIOGRAPHIC STANDARDS OFFICE. Name/title authority files in the B.L.: a discussion paper. Unpublished, 1976.
- 4 LINFORD, J. Draft EUDISED format in *Council of Europe: Documentation Centre for Education in Europe*. EUDISED: standards, format, character representation. The Council, 1973, p. 45-90.
- 5 INFORMATICS CONFERENCE. Lancaster, 1976. Proceedings not published.
- 6 NOERR, P. L. and WHITE, M. C. MERLIN: design of a national bibliographic database in *Systems for large data bases: Proceedings of a 2nd International Conference, Brussels, 8th-10th September 1976*, LOCKEMANN, P. C. and NEWHOLD, E. J. (eds). North Holland, 1977.
- 7 ROSS, J. and ROYAN, B. MERLIN: A new computer system for the British Library *Program*, 10 (3), p. 95-102.