

CONVERSION OF THE BRITISH LIBRARY'S MAP CATALOGUES: THE KEYS TO SUCCESS

By Tony Campbell

In this paper I shall outline the reasons, both historical and technical, which led to the British Library's choice of retroconversion method. I shall describe our progress to date and our future plans, with particular attention focused on the "Data Conversion Specification". Even when I descend into detail, the intention is to describe an experience which might be shared by others following in our footsteps. It is not my concern - you will be relieved to hear - to itemise every peculiarity of the BL's catalogues and our ingenious solutions for dealing with them.

Unlike many long-established institutions, the British Library has been able to produce printed map catalogues. The last of these was published only fifteen years ago. Why do these need to be converted?

I hope that so distinguished an international gathering as this does not require a serious answer to that question. The benefits of automated access should be now be too well established to need explanation or justification. In the British Library's case, there were the added advantages that would flow from merging separate catalogues and the improvements made to authority controlled headings. The overriding impetus, however, came from the automated needs of the new BL building at St Pancras, into which we shall be moving in a few years time.

The question then, was not whether we should convert but how. To understand why there was only one realistic procedure it is necessary to review briefly the history of the BL map catalogues. Many of you will know The British Museum catalogue of printed maps, charts and plans, published in 15 large red volumes in 1967. What might seem surprising is the comment on the verso of its title, 'first published 1885'. How is this possible when so many of the entries are post-1885? The explanation is that the catalogue was first published in 1885 (in 2 volumes) followed thereafter by printed annual supplements. To update the master set of the catalogue, these supplements were chopped up and their entries distributed into guard-book volumes. The preparatory work for the 15-volume publication therefore involved tidying up the sequence so that it could be lithographically reproduced. Whereas the 1885 catalogue had been typographically uniform, the 1967 catalogue was a *mélange* of type styles and variant layouts. In addition, some en-

tries were almost too faint to be read and there had been a number of handwritten annotations.

Thus, when the possibility of conversion was first seriously discussed in 1989, optical character recognition was quickly ruled out. This meant the catalogue would have to be keyed. Because of its size - about 210.000 main (i.e. geographical) records - the operation would have to be done outside the BL by a commercial bureau. As the anticipated costs were more than the BL was prepared to meet, it was also necessary to offset the keyboarding costs against exclusive publishing rights. The publisher and the conversion bureau were therefore sought through a two-part tendering process.

The complexities and inconsistencies in the catalogue were such that a detailed Specification had to be drawn up, giving precise instructions about identifying and converting every type of data element involved. Two successive managers, James Elliot and Geoff Hutt, worked on producing the 'Data Conversion Specification' (a 93-page document). Preparation took well over a year, followed by a further six months for completion of the tender process. The Tender Board, comprising a range of expertise from across the Library, completed its work before new European Commission rules came into force. The requirement that publicly-funded tenders must be advertised throughout the EEC will significantly lengthen such operations in future. Finally, in April 1992, we were able to sign a contract with Research Publications International, awarding them the right to publish the catalogue on CD-ROM (with a royalty payable to the BL). They, in turn, subcontracted the keying to Access Innovations (of New Mexico). Keying of the 15 volumes, and the similarly organised 10-years supplement covering the period up to 1974, was scheduled to have been completed within about two years. In the event, the test phase took longer than expected but the project is now running smoothly.

Publication of the first CD-ROM is currently planned for 1995 and work has already started on the interface design. To help devise appropriate screen menus and search facilities, an international editorial board has been appointed, under my chairmanship. Those members who were available met for the first time at the International Conference on the History of Cartography in Chicago in June. At the LIBER conference, here in Barcelona last September, a questionnaire was circulated to obtain librarians' views on the type of use anticipated and details of the search strategies required. A reworked version of the questionnaire will be sent out later to those in the recently published Third Edition of the World Directory of Map Collections - for the which the IFLA Geography and Map Libraries Section thanks Lorraine Dubreuil.

I will now talk in more detail about the Specification and how this controls the transformation of the printed text into machine readable data. The complexity is the result of the same long history to which I have already referred.

Had all the catalogue descriptions been created according to consistent rules, or had the body of each record (rather than just the geographical headings) been edited before publication of the 15-volume catalogue, the Specification would have been slim and straightforward. The reality was very different. Records created since World War II, while not conforming to the Anglo American Cataloguing Rules we now use, nevertheless contain most of the necessary elements. Material described pre-War, however, was treated according to the cataloguing rules current at the time. There have been several of these, stretching back to the minimalist entries for King George III's Topographical Collection (K. Top.), received by the then British Museum in 1828.

For the bibliographic archaeologist it may be an interesting exercise to reconstruct the rules and the periods during which they operated. As far as the Retroconversion was concerned, though, it meant that the conversion bureau had to be systematically warned of all the situations they might meet so that the coding staff would correctly identify the fields and subfields.

The variations that are signalled in the Specification are of three types: elements that are not always present in the entry, elements that may be found in different relative positions, and variations of typography and punctuation. The most rudimentary entries of all - those for K. Top. - may occasionally contain no more than a title. Most of the other records have at least the basic ingredients of a catalogue entry but may lack one or all of the following: place of publication, publisher, size, scale, engraver's name or notes. It is worth pointing out that it has never been BL policy to include coordinates or to specify the projection. Nor do we catalogue map series at sheet level.

The second unpredictability concerned the relative positions of elements in a record. The most important inconsistency relates to the place of publication and publisher, which were inverted in earlier cataloguing rules. Typeface variation is the least noticeable of the inconsistencies that have to be described. Precisely for that reason is proved the most complex to analyse. If a heading type was to be recognised by the coding staff because it was in bold or italics, then it was essential to note that sometimes the same kind of heading might be in capital letters. The same applied to punctuation and to prescribed combinations of typefaces. The many variations produced by cataloguers working over a period of more than 150 years, and the wide range in the types of original material being described, produced a large number of distinct situations to be treated in the Specification.

At the end of the conversion, the data is delivered in a simplified UK MARC format. However, the coding staff do not need to know the UK MARC tags nor the subfield structure. Their task is to identify correctly each bibliographical element (title, scale, the three imprint elements, physical description, shelfmark, etc.) and assign the appropriate letter code to each. These initial codes are later converted automatically to the MARC format. Coding is in many ways more important than keying. Data allocated to the wrong field may well not appear when the completed file is later searched.

Once a section of the catalogue has been coded, it is passed to the keyboarders for data capture. The Specification contains detailed rules of capture to ensure that consistency is imposed on typography, punctuation and spacing. Bold, italics and small capitals are not used - although colour - coding and highlighting on screen can serve similar purposes. Thus, many of the instructions concern typeface simplification. Others explain how to deal with name added entries, cross-references and non-Roman scripts. For name added entries, just the heading and shelfmark are captured. By matching these shelfmarks against those in the main geographical entries, the name headings will be automatically inserted into the appropriate main entry at a later date. With the exception of cyrillic, non-Roman scripts have not been captured; instead they have been flagged for future transliteration by language specialists.

The most complex issue concerns the catalogue's use of modular rather than the more usual unit records. All editions, facsimiles, impressions etc. of a work are listed perfunctorily as part of a single record, rather than as separate unit records. The printed catalogue is arranged, however, by the date of the first edition. An atlas of 1934, is an admittedly extreme example, appears under 1816 in the sequence, since it is the 10th edition of a work of that earlier date. In conversion, each edition had to be given a proper identity of its own. They have therefore been captured as a distinct subrecord within the main record for the earliest edition. This enables each edition to be searched by its own date and retrieved separately, for instance in a geographical query, while also continuing to allow all editions of a map to be retrieved together. Thus, in the example given, anyone searching for atlases from the 1930s would be offered that 10th edition of 1934, once the catalogue is converted.

Other keyboarding operations add elements or repeat them. While dropped headings are appropriate for a printed catalogue, each automated record must be complete in itself and so the headings are repeated. For ease of searching, the representative fraction, which may occur in the title or the verbal scale statement, is written into a special field as a pure number (i.e. "Scale 1:63,360" becomes simply

"63360"). The date of publication is copied into a searching field in a similar way. The remainder of the Specification is taken up with sample pages from the catalogue showing the correct coding, and instructions on conversion to MARC format and output to magnetic tape. A long appendix describes the character sets used in the conversion - the same specially extended range of characters devised for the conversion of the BL's general book catalogue (BLC).

As a general rule, the catalogue was converted unchanged. There was no attempt to look for errors and inconsistencies and no material was recatalogued. Given that the equivalent of less than two members of staff are available to work on the project, this decision was unavoidable.

However, it did prove possible to carry out some pre-editing, mostly to make the geographical headings more consistent and useful.

This in turn will make possible the automatic insertion into each record of a code (lifted from the Dewey Decimal Classification) representing the geographical area covered. The codes themselves are hierarchical, and will therefore enable a region, and all the places within that same region, to be retrieved in a single search. The other pre-editing operations involved transferring hand-written corrections from four working sets of the catalogue to the conversion master, and reinstating some subject-headings that had been suppressed during the 1960s editing operations.

My final remarks describe current progress and future plans. After some delays, the first volume has now been accepted. A random 10% sample (selected by the BL) was proof-read and the errors were found to fall below the permitted maximums: 1.5 errors per thousand characters for codes and the more significant fields and 3 per thousand for the rest. If a sample fails, the entire volume has to be corrected and a new sample proof-read.

We now expect to deal the volumes at the rate of one a month. Following that, attention will turn to keying the 3-volume catalogue of manuscript maps, produced in the period 1844-61. Those two new files will then be merged with the current file containing material catalogued since 1975, which has been in machine readable form since 1988.

A range of automated concordances will be run against the converted data. Among other enhancements, these will impose consistency on the shelfmarks, supply normalised place of publication and the country of publication code, and add the geographical coding already mentioned. Like much else, these operations involve close collaboration between the Map Library and the BL's computer departments. Map Library staff prepare the concordance lists, usually in the form of

paired columns running to several hundred entries each, in either word-processed or database format. The computer section then provides the necessary programming to apply this data to the converted records, just as they had earlier been responsible for the detailed analysis of our requirements. This paper has, I hope, spared you the highly technical aspects of our retroconversion. Nevertheless, the contribution of specialist computer expertise to a project such as this is considerable.

The resultant file will be published as the first edition of a CD-ROM. It will also be made available on the OPAC that is now being developed for the new BL building. Later editions, which will punctuate the 15-year partnership agreement with Research Publications, will incorporate the results of manual editing and add further cartographic descriptions: records from the BL book files, entries in general BL catalogues, handwritten catalogues of specialised map collections, and material never previously catalogued. This last category includes the descriptions of atlas maps on which the independent researcher Rodney Shirley is engaged. The ultimate aim is for comprehensive and unified access to the 300,000 identified cartographic entities spread throughout the various BL departments.

The end of the story, perhaps, is the contribution such a file might make to international record sharing. I shall be talking about such possibilities in my paper to the Geography and Map Libraries Section meeting on Wednesday, which I hope many of you will be able to attend.

Tony Campbell
British Library
Map Library
Great Russel Street
London WC1B 3DG
UK