

APPENDIX I

THE BIBLIOGRAPHIC/MARC PROCESSING SYSTEM

The Bibliographic/MARC Processing System (BPS) was developed to create, maintain, and process the machine-readable catalog of the Library of the UNC Carolina Population Center*, and to produce various catalog-based products including lists of acquisitions, lists of publishers, authority lists, subject bibliographies (with name and subject indexes), and a microfiche catalog and supplements. Due to its considerable flexibility, the system readily accommodates many other types of information, and has since been employed by various UNC organizations for cataloging and retrieving press releases, machine-readable datasets, computer programs, audio-visual materials, correspondence files, musical works, and public opinion survey items. The BPS programs have limited tabulation capabilities as well, and have been used to create and analyze a file of manpower training records for the U.S. Agency of International Development.

The MARC Record Format

The record structure used by the BPS programs is based on the MARC (an acronym for "Machine Readable Catalog") record format developed at the Library of Congress for the specific purpose of encoding bibliographic data in machine readable form. This record format has become widely accepted as a standard format for the processing of bibliographic information. The essential characteristic of MARC-format records is that they can accommodate a varying number of variable-length data items, hence their great flexibility and broad applicability.

Each MARC record begins with a 92-byte leader, which contains the record identifying number (called the "did"--document identifier), the creation date, and other fixed-length data and control information, including a count of the number of variable-length data fields contained in a later part of the record. A record directory contains an entry for each variable-length field, giving its location in the record, its length, and an abbreviation identifying the type of information it contains. Additional descriptive information is found in the first bytes of the variable-length field itself.

Each field is associated with a three-letter code called a "tag." For example, in the CPC Library catalog, "tit" refers to the document title, "cal" to its call number, "imp" to its imprint. In the USAID Manpower Training Record file, "nme" is the individual participant's name, "nat" his or her nationality, and "spn" the sponsor of the training.

*With the support of a grant from the U.S. Agency for International Development

Though tags are generally defined when an application is being structured for use of the Bibliographic/MARC Processing System, new tags may be defined any time a need for them is perceived.

Any individual record may contain from one to one hundred tagged fields. A record may also contain more than one field of the same type. For example, a document might be associated with several primary subject terms ("sup"), several primary geographical area references ("gap"), etc. Multiple occurrences of the same tag in a record are differentiated by tag sequence or "site" numbers.

Each variable field may be further subdivided into variable length subfields. Each subfield is associated with a letter of the alphabet (the "subfield code"). A field may consist of a single subfield or as many as twelve subfields. A CPC Library example would be the imprint (tag "imp"), where the place of publication is subfield "a," the publisher subfield "b," and the year of publication subfield "c."

Finally, tagged fields of the same type may be further distinguished by the use of a tag suffix or "indicator." Indicators serve to denote some minor differences between two fields that in other respects are to be treated identically and which have therefore been assigned the same tag.

Processing Capabilities

Creation, correction, and updating of the MARC-format master file are accomplished through the use of the Update programs. New records are added and existing records are changed in exactly the same manner as the original master file is created.

System capabilities for creating new records or changing existing records enable the user to:

- Add, delete, or replace an entire MARC-record.
- Add, delete, or replace any variable field or fields in a record.
- Change any text string in any variable field or fields.
- Change subfield codes or indicators, without reentering field data.
- Enter the same change for a series of consecutively numbered records, for a range of contiguous records whose "did" numbers are not consecutive, or for multiple records that are not contiguous.
- Print the entire record following any change, or print only the fields affected.

- Display any other record or group of records, without changing them.
- Make systematic changes to one or more variable fields in all records in the file.
- Process transaction lines generated by the Print program (BPSPT).

A range of facilities for generating printed reports, catalog cards, and machine-readable files is available. Most types of output are produced by general purpose selection, sort, and print programs, according to instructions coded into option statements by the user. Certain specialized forms of output, such as a key-word-in-context (KWIC) listing, are produced using programs designed specifically for those tasks.

System capabilities for retrieving and displaying information from the data base include:

- Selection of records based on the presence, absence, or content of one or more variable or fixed fields, according to a selection rule containing logical operators "AND" and "OR."
- Sorting of records by one or more variable and fixed fields.
- Special handling of missing data fields, including provision for placing such records at the beginning or end, for specifying an alternate field for sorting, and for excluding such records from further processing.
- Execution-time specification of fields to be printed, including modification of terminal punctuation and insertion of user-supplied captions.
- Generation of report titles and page headings with (optional) page numbers.
- Choice of single or multi-column format, with provision for keeping data for records or groups of records from being split across columns or pages.
- Specification of column width, page size, indentions, and paragraph spacing, if default values are not desired.
- Underscoring, dark printing (overstriking), and/or converting to upper case of data from selected fields or subfields.
- Tabulation of records within categories and subcategories for the preparation of summary reports.
- Production of indexes to printed reports.

Principal Computer Programs

The principal computer programs that comprise the Bibliographic/MARC Processing System are:

1. Data Edit Program (BPSDATA) -- checks the document identifying numbers of new records and of corrections to existing records, and translates these numbers to an internal-format number suitable for sorting by the IBM Sort program. Program BPSDATA also has several facilities for expediting the entry of identical changes or additions to multiple records. The records produced by BPSDATA (called transaction records) are sorted by the IBM Sort program and then read by the Update program.
2. Update Program (BPSUPDT) -- checks transactions for a variety of error conditions, adds, changes, deletes, prints, and replaces records; adds, changes, deletes, and replaces fields within records; produces a proof sheet listing errors detected and the current contents of records or fields changed.

Extended features include a record selection capability (as an efficient procedure for selecting records whose "did" numbers are known), an extraction capability (to produce an abbreviated version of the file for special uses), and an edit capability (to change all occurrences of a specific data item throughout a range of records or the whole master file--for example, changing the name of an organization following a merger).

3. Option Statement Preprocessor (QUERY) -- reads option statements describing specifications and requirements for selection, re-ordering and display, invokes standard option sequences from an option library, and passes option statements to the programs below.
4. Free Text Retrieval Program (BPSRET) -- selects records from the master file according to one or more sets of search criteria. Several independent searches can be carried out simultaneously. Record selection can be based on the presence or absence of particular variable fields, or on the content of variable or fixed field data. Compound selection rules can be formed from individual selection criteria and logical operators.
5. Sort Key Edit Program (BPSSKED) -- prepares master file records (after selection as above or the entire file) for re-ordering by constructing a sort key according to parameters entered at execution time. The sort key is appended to the front of the MARC record to enable the file to be processed by the standard IBM Sort/Merge program.

The Sort Key Edit Program also generates additional copies of master file records where the same information is to appear at several places in a printed report (for example, a subject index

would list the record under each of its subject terms.) The facilities of this program and the following one enable master file information to be presented in ways very different from its original record organization.

6. Print Program (BPSPT) -- prints data from master file records, with or without sort keys, according to specifications entered at execution time. Great flexibility is afforded for the content and positioning of report titles, page headings, and record data. Summary reports, tabulations, and indexes to previous reports may also be produced with this and the previous program.

Written for RFP response
on Hospital Organization Research

4/28/80