

MINISIS - A MINICOMPUTER BASED BIBLIOGRAPHIC SYSTEM

Anne Edwards and Kate Wild
International Development Research Centre
Association of Information and Dissemination
Centers Spring Meeting
June 3rd - 5th, 1979



The International Development Research Centre was established to support development research in Third World countries in the fields of: Agriculture, Food and Nutrition; Health Sciences; Social Sciences; and Information Sciences.

The IDRC Library, which is part of the Information Sciences Division, was conceived not only as an information resource for IDRC staff, project staff and Canadians with an interest in development, but also as a testing ground for various techniques and methodologies that could be applied in information science projects in developing countries.

The decision to computerise the library was thus taken in 1973 not only to provide access to the library's specialised collection in the field of development but also to create a basis of experience which could be drawn upon when advising Third world countries on the establishment of bibliographic information systems.

The system that was chosen in 1973 was ISIS (Integrated Set of Information Systems) developed at the International Labour Office in Geneva. ISIS runs on large IBM computers and IDRC operated ISIS through a service bureau for several years. Data entry and retrieval were carried out on-line at IDRC. The production of indexes, lists, purchase orders - any printed output except search results printed on a hard copy terminal - and the updating and management of files were batch operations carried out at the service bureau where we submitted the appropriate job control cards.

During the first three years that we had ISIS in operation, we created our own data base of some 15000 records and provided search services on this as well as on the data bases of ILO, FAO, and UNIDO. We also built a small data base relating to health care delivery systems in developing countries.

Apart from the inconvenience of the service bureau arrangement, by 1975 it was clear that operating through a service bureau on a large computer was a very expensive proposition (in the order of \$120 000 to \$140 000 per year). We decided to investigate the possibility of acquiring a minicomputer to run the library's operations, to maintain the specialised data bases we were already making available, and to develop a relatively inexpensive and reliable package of both hardware and software that could be transferred to developing country institutions.

The new system would be used to build the IDRC library data base (from acquisitions through to retrieval and production of indexes); to build the DEVSIS file (a data base to test the design of an international system to provide information about economic and social development); and to build the SALUS file (a data base recording information about low-cost rural health care and health manpower training). From both DEVSIS and SALUS regular annotated bibliographies would be produced.

After an evaluation of minicomputer manufacturers and their products, and after contacting other institutions where information systems for minicomputers were being developed, the Hewlett-Packard 3000 Series II was selected and thoughts turned to the design of the system. There were several alternatives. The ISIS programs could have been recoded for the HP 3000, but this solution would not have taken advantage of the special features of the HP 3000, nor would it have represented a significant saving in man years for analysis and programming. The data base management package (IMAGE) developed by Hewlett-Packard for the 3000 could have been adopted, but it was found to be unsuitable for a bibliographic system where imbedded keys and sub-fields were required.

We decided to design a new system, but one that would perform all the functions of ISIS and that would enable us to continue to exchange information with ISIS users. We knew that, from the system designers point of view, the system should meet certain requirements. It should be:

- general purpose: it should not be restricted to the processing of bibliographic information;
- modular: it should be possible to debug, maintain and extend segments of the system without impact on the system as a whole;
- independent: the application programs should be independent of the data base management system so that in a multi-user environment it would be possible to maintain a common basic system.

From the users' point of view it should meet other requirements. It should:

- handle data in a number of physical forms;
- be simple enough to understand so that the user would have control over the tasks he needed to perform;
- provide a wide variety of outputs.

From the manager's point of view, still other requirements needed to be met. The system should:

- be cost effective so that the cost of hardware could be written off over a specified period of time;
- be able to accept output from other systems so that the particular needs of users could be satisfied;
- be compatible with other international information systems for which it could be used as an input and output system.

These then were our guiding principles.

System development started in June 1976. Consultations were carried out with the library on its needs and regular meetings were held between librarians and the computer group.

Coinciding with the design of the new system, the library made a move towards promoting a standardised bibliographic exchange format by adopting the recommendations of the "Reference manual for machine-readable bibliographic descriptions" (UNISIST/ICSU-AB Working Group on Bibliographic Descriptions, 1974) for formatting its bibliographic records. The DEVSIS Study Team had already recommended this format for creating a world-wide data base of literature in the development sciences.

MINISIS AND THE USER

To the user the system is, in some respects similar to ISIS. It performs all the same library management and information retrieval functions that the IDRC version of ISIS performed. But it also allows different users to view the data base in the way most appropriate to their particular interests. Thus the acquisitions staff see, besides selected bibliographic fields, all those fields which relate to ordering, whereas the cataloguers see all the bibliographic fields but none of the fields relating to ordering procedures.

Data Base Building

Very briefly, the creation of the library data base can be described as follows:

Information entered on line by the acquisitions section is used to generate computer printed purchase orders on specially designed forms. When the ordered item is received, the same record that generated the order is accessed from the cataloguing point of view in order to update the cataloguing data and add the abstract. Records are written in English, French or Spanish (the three working languages of the IDRC) according to the language of the document. Once the cataloguing data has been proofread and pronounced "clean", it is "released" by the data base manager for retrieval. Subject access to the collection is via on line searching of descriptors assigned to each item from the trilingual "Macrothesaurus for information processing in the field of economic and social development" (Organisation for Economic Co-operation and Development, 1978). Prior to the publication of this second edition of the Macrothesaurus, the IDRC version of the first edition was used. The library presently has a data base of some 25 000 records comprising records for monographs, analytics and serials.

There are five automated authority files linked to both the library data base and the DEVSIS data base.

(1) The Authority of Institution Names

This file holds the name, location and the ISO two-letter country codes (ISO-3166. Code for the representation of names of countries, 1974) of institutions (whether identified as a corporate author or an affiliation), that correspond, for each bibliographic record, with the "place where the work done". The file includes "see" references, other language versions, and broader-term and narrower-term relations. A six-digit code entered in the corporate author or affiliation field of the bibliographic record constitutes the link with the appropriate entry in the Authority File of Institution Names. These codes can be used in bibliographic searching in combination with descriptors in order to retrieve items on a particular subject produced by a particular institution.

(2) The Thesaurus

This is fully trilingual. There are English, French and Spanish descriptors for each concept. A search conducted in one language will retrieve all relevant items, whatever the language in which the descriptors were originally assigned. The thesaurus structure is built into the on-line searching facility so that, for instance, all the narrower terms of a given descriptor can be searched.

(3) The Vendor Authority

Here are maintained the name and address of each book supplier used by the library. It allows names and addresses to be printed out on purchase orders when a four-letter code is entered in the vendor field of records.

(4) The Non-Vendor

Four-letter codes are expanded to appropriate trilingual messages on the bottom of the purchase order when the requested item can be obtained free or on exchange.

(5) A system-wide trilingual stop word list

This eliminates "noise" words (prepositions, pronouns, conjunctions, etc.) on request.

The system is fully interactive. Aspects of computerised library operations which formerly needed to be carried out by computer staff are now carried out by library staff. This is one of the main reasons why we believe the system to be more user-oriented than ISIS. The terminal operator orders her own proof-lists on line; global changes to the data base can be made on-line; printing of purchase orders is commanded on-line. The data base manager creates indexes by selecting the fields on which the index will be created and designing the print format, and then either prints them out immediately or streams them to run later. The print format and index specifications can be stored and called upon at any time to produce further printed indexes or to prepare output tapes for COM processing. Many users may use the same data base at the same time to do many different things.

The system handles all French and Spanish diacritics using the options available within the ISO 7-bit coded character set (ISO-646, 7-bit coded character set for information processing exchange, 1973) and it has full upper and lower case capability.

Two processors are available to users for data base building:-

- ENTRY processes new records. ISN's (Internal Sequence Numbers = record numbers) are generated automatically by the system for those data bases requiring access by ISN. Fields are prompted in language easily understood by the inputter. ENTRY incorporates two very useful features:
 - any field can be automatically checked to ensure that the item being entered is not a duplicate, thus eliminating a great deal of clerical checking;
 - fields requiring authorities can be validated against their respective authority files. Validation can be carried out either on-line or in batch.

Dual entry, i.e., entry into more than one file at a time, is another important feature of this processor. For example, a new record can be created in an authority file at the same time as a new bibliographic record is being created.

- MODIFY processes changes to records already in the system. A record is accessed in MODIFY either by

specifying its ISN or through a query on any desired field. Fields to be modified are accessed either by field tag (consisting of one letter and three digits) or by short mnemonic field name. Fields can be added, deleted, replaced, or changed, and information in one field can be transferred to another field. In MODIFY, global changes (i.e. the same change to the same field across the data base) can also be carried out using identical methods to those used to modify a single record.

Two other processors are used for output:

- INDEX sorts records to produce, for example, KWIC or KWOC indexes, alphabetical author or title listings, or a shelf list. INDEX can handle the UDC sort sequence (the Universal Decimal Classification is used in the IDPC Library), and can also handle the Spanish-language alphabetic sort sequence.
- PRINT specifies the format for printing each record (order and arrangement of fields; insertion of literals and punctuation) and specifies the page lay-out (number of records per page; whether printed in columns, tables, etc.).

RETRIEVAL

It seems we have spent a lot of time talking about the system before we come at last to the subject which is the theme of this session - Retrieval.

The processor which carries out retrieval is QUERY.

Any field in the data base can be used for retrieval using either inverted file techniques or free text searching on a previously selected sub-set. Two kinds of search strategy are available - entering each operand together with its logical relationship to the other operands, or entering each operand separately and combining them in a subsequent operation. Operators are "AND", "OR", "AND NOT" and "EOR" (either one term or the other but not both). Postings are displayed at each step of a search.

An easily understood set of commands can be displayed on the screen at any time during a search by typing "Help".

The command "Browse" allows the items found in a search to be displayed on the screen. The command "List offline" causes the results to be printed. Fields displayed or printed depend entirely on the wishes of the user. Any number of print formats can be pre-defined (using the PRINT function) and the command "Format (print format name)" can be given at any time during the search to change the format used for displaying or printing. The user can "Keep" a

search formulation for use at a later time, or "Save" the set of records retrieved for use in the INDEX function.

Searching on a descriptor in one language automatically retrieves items indexed in all three languages. However, a simple command allows the user to specify that only one or two of the three languages be searched. The user may specify that separate postings be displayed for each language.

Other commands allow the current search formulation to be displayed or a list of inverted fields to be displayed.

MINISIS supports the use of thesaurus structure for subject retrieval. The Thesaurus is matched against descriptors in the data base to form an inverted file. There are Broader Term, Narrower Term and Related Term links between the descriptors. "Any Tables" have been constructed to group together terms which are often associated with each other in searching, for example, all countries within a particular region or continent.

A command in the QUERY function also allows the user to perform arithmetic operations on fields containing numeric values. This feature is used in the Library for calculating, for example, the amount of money committed to unfilled orders, or for calculating the amount of money spent over a particular period. It is also used by Administration when querying the IDRC Project Information System (PINS) data base.

However, for producing printed reports based on fields containing numeric values, the COMPUTE processor is used. It generates pseudo-fields containing the results of arithmetic operations which can be sent to the PRINT processor.

We said earlier that we were prompted to convert to a mini-computer system for reasons of cost and convenience, and in order to develop a viable small-scale information management and retrieval system for developing country institutions.

Since we acquired the Series II, the Series III has come on the market, providing more capabilities at less cost. MINISIS will run just as well, if not better, on a Series III and will also run on a Series 33.

The cost of purchasing a Series 33 with a 50 megabyte disc (but no tape drive) is \$70 000. The 33 provides a maximum of 1 megabyte of memory and 480 megabytes of storage. The Series III has a maximum of 2 megabytes of memory and 960 megabytes of storage. Its price, including a 50 megabyte disc and tape drive, is \$115 000. The cost of acquiring a Series 33 and peripherals for an organisation the size of IDRC would be around \$135 000 and the cost of a Series III for the same size of organisation would be

around \$180 000. This represents a smaller investment than was made by IDRC in 1976 to acquire the Series II. Costs are going down!

In the final analysis, obviously, the system is less important than the information it retrieves. At IDRC we can now provide access from the following files:

- IDRC - around 25 000 records giving bibliographic description and subject analysis of the library's holdings in subject areas relating to development.
- DEVSTIS - 2000 records representing current literature on development produced in Canada, the Federal Republic of Germany, Pakistan and the Philippines.
- SALUS - 5000 records giving bibliographic description and long informative abstracts of documents from all over the world on low-cost rural health care and health manpower planning.
- FAO - 40 000 records representing the holdings of the FAO Library.
- ILO - 90 000 records representing the holdings of the ILO Library.
- UNIDO - 7000 records representing the holdings of the UNIDO Library.
- UNESCO - 16 000 records relating to Unesco documents and publications.

We cannot at the moment make all these files available simultaneously, but we hope to be able to do so in the near future when we add further disc capacity to our HP.

In some of our files, retrieval capabilities are quite original. In the DEVSTIS file we can retrieve, for example, documents on agricultural projects in any African country with an investment value of over \$100 000.

We have been approached to provide MINISIS software to a number of institutions in other countries.

The International Labour Office in Geneva will use MINISIS not only to provide data base management and retrieval on the Library's large data base but also to manage their mailing list, the logging in of incoming mail and telexes, the acquisition of equipment for projects, their project information file, and their candidates file.

The Agricultural University at Wageningen in the Netherlands will use MINISIS to run the large University Library.

The Centre national de documentation agricole in Tunisia will use the system for data base building and retrieval and plans to load AGRIS tapes for retrieval.

INION in the USSR will use the system for, amongst other things, building a data base in the framework of DEVSIS.

In Canada, the Sport Information Resource Centre uses MINISIS to manage a bibliographic data base of 50 000 records; and the Department of Health and Welfare is using the system to build a data base of reports on safety and health with regard to chemicals. Health and Welfare had previously used a service bureau which they found extremely costly and limited in capability.

Some of these institutions not only intend to use MINISIS but also wish to participate in cooperative projects with IDRC. They wish to use MINISIS to help build data bases of literature in agriculture and in economic and social development. Much of this information is unpublished and has never before been recorded in any library or information system. By creating computerised files and making them available to international systems, such information will become available for the first time.

Thus, the main reason we embarked on the MINISIS program was to encourage the recording and sharing of valuable unpublished literature. Our decision appears to have been justified by the results.