DANELI 30136

MINISIS - A MINICOMPUTER-BASED BIBLIOGRAPHIC SYSTEM

Paper prepared by Faye A. Daneliuk (Project Manager, Minicomputer
Project) and Anne Edwards (Systems Librarian) of the
International Development Research Centre, Ottawa, for the West
Canada ASIS tenth annual meeting, Winnipeg, September 27 to 29,
1978.

Abstract

The International Development Research Centre (IDRC) in Ottawa
has developed a data base management system, oriented to the
handling of bibliographic information, on a Hewlett-Packard 3000
Series II. The paper describes the environment for which the
system was developed; management requirements for the choice of
hardware and the development of software; and how the system is
used in the IDRC.

The International Development Research Centre both carries out
and funds research for development in several fields, including the
field of Information Science, for which, we believe, it is unique. So
the decision to computerize the IDRC Library in 1973 was made, not
only to provide a means of accessing the library's very specialized
collection of literature on the subject of development, but also to
create a basis of experience for establishing computerized
bibliographic systems in developing countries.

The system that was chosen in 1973 was ISIS (Integrated Set of
Information Systems) developed for on-line applications in the Library
of the International Labour Office in Geneva. IDRC thus became part
of a "network" of ISIS users around the world who made available to
each other developments or refinements in the system. IDRC has helped
to install ISIS in several developing countries including Chile,
Kuwait, Costa Rica and the Philippines.

ISIS depends on the use of a large IBM computer and IDRC ran ISIS
through a service bureau. Operating through a service bureau is
particularly expensive, but operation of this type of on-line system
on any large computer is expensive. In 1975 it was decided to
investigate the possibility of acquiring an in-house minicomputer, to
run the library's operations, to maintain a number of specialized data
bases, and to develop an inexpensive and reliable package of both
hardware and software that would be transportable to developing
countries.
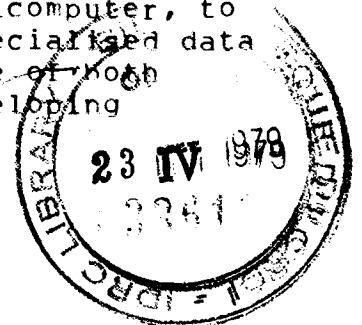
The new system would be used to produce Devindex Canada, an abstract bulletin of Canadian literature on the subject of development which the Information Sciences Division publishes in the context of DEVSIS (Development Sciences Information System), and it would also be installed in institutes in developing countries participating in DEVSIS. Moreover, the Information Science Division helps fund certain AGRIS (the Agricultural Information System of the United Nations Food and Agriculture Organization) participating centres and the package would need to be suitable for processing AGRIS records in such centres.

Accordingly, in April 1975, a consultant was employed by the Information Sciences Division to study the feasibility of implementing a bibliographic information system on a minicomputer. The resulting report, which took into account such factors as costs and software development, was positive in its conclusions and justified more intensive investigation on the part of IDRC.

In the following eight months a critical evaluation was made of all "stable" minicomputer manufacturers in the market and of their products. The cost of the equipment, although an important factor, was not the only factor taken into consideration. Because it was intended to do the software development in-house, the extensiveness and reliability of the manufacturer's software was studied in great detail. The end-product (the information system software) was being developed not only for use within IDRC, but also for use in areas where an inexpensive, reliable facility rather than a large-capacity one would be the requirement. This meant that we had to have some assurance that the manufacturer would remain in business for some time. The manufacturer also had to provide some form of service for his equipment in Africa, Southeast Asia, and South and Central America, for we hoped to make the software available to institutions in those areas. (This latter requirement was not met by any of the minicomputer manufacturers!)

At the same time, other institutions where information systems were being developed for minicomputers were contacted. In some of these institutions, a machine had been selected to run a dedicated system, and our first inclination was to adopt the same procedure. It can be easily understood that if a machine is dedicated to one application then the manufacturer's software need not be overly sophisticated. However, during this period of evaluation, an on-going dialogue was taking place between Information Sciences and other divisions within the Centre. It was decided that the acquisition of a somewhat more sophisticated computer could be of great benefit to the Centre itself. This led to a narrowing of the field of potential suppliers.

In early 1976, a project proposal dealing with both the acquisition of the minicomputer system and software development, was

presented to the Board of Governors by the Information Sciences
Division. The project, as approved by the Board, specified three
major reasons for a minicomputer:

     (1) to reduce operational costs of running ISIS at IDRC (this
was a significant factor - running costs at a service bureau can be
extremely high);

     (2) to define an optimum cost-benefit minicomputer installation
that could be offered, complete with programs, for AGRIS (Agricultural
Information System of the Food and Agricultural Organization of the
United Nations)/DEVSIS (Development Sciences Information System)/ISIS
activities at national centres in developing countries;

     (3) to provide a basis for a Canadian input/output centre to a
future international network for development information (DEVSIS).

     The three-member "computer group" within the Information Sciences
Division drew up a tender that stressed three characteristics: (1) the
power of the operating system; (2) the reliability and availability of
the manufacturer's software; and (3) the potential of the machine to
handle the job mix. The same group also designed and conducted
benchmark tests that emphasized the three critical specifications in
the tender. Hewlett-Packard won the tender with their 3000 system.
Although the HP-3000 had some short-comings, which are only now being
corrected, some fine hardware features (stack architecture) and their
sound, integrated operating system (MPE - Multiprogramming Executive)
helped them to win the tender. In August 1976 our equipment was
delivered and development began on the software.

     Once the computer was selected, thoughts turned to the system
design. A final decision had yet to be made that we were indeed going
to develop new software for our information system. Other
alternatives did exist. We could simply recode the ISIS programs for
the HP-3000 or we could adopt the data-base management package (IMAGE)
developed by Hewlett-Packard for the 3000. The first alternative,
although providing a quick and easy solution to a system-design
problem, would have proven unrealistic because it could not have taken
advantage of the special features of the HP-3000. The second
alternative demanded more consideration but was finally rejected for
reasons familiar to those who have worked with bibliographic systems:
(1) no capability for handling true variable-length records; (2) no
capability for handling variable length, variably occurring fields;
(3) no capability for handling subfields; (4) no capability for
supporting keys embedded in text; and (5) no capability for handling
long descriptive abstracts. A new design was definitely called for!

As a first step in designing the system, a number of guiding principles were adopted:

(1) general applicability - the system should be as general-purpose as possible;

(2) modularity - the system should be totally modular to promote ease of maintenance and extension;

(3) independence - the applications functions should be independent of the data-base management functions;

(4) user considerations - (a) the system should be flexible in that it should be capable of handling data in almost any physical form; (b) the system should be simple to understand so that it could be implemented and used with minimum effort; (c) a user-attractive language should be provided so that users could really be users; and (d) the system should be able to provide a wide variety of outputs;

(5) mission orientation - (a) the system should have the capacity to accept outputs from other information systems; (b) the system should be viable within a small organization; and (c) the system should be compatible with other international systems, specifically ISIS and AGRIS;

(6) cost-effectiveness - the basic system should be in operation by December 1977 so that we could dissociate ourselves from the service bureau where we had been running ISIS.

In designing a system, a set of guiding principles, though very important, is not sufficient. Also required is a theoretical frame-work around which to build the system. This framework provides a coherence that otherwise would be difficult to realize. Careful study was given to the three prominent data-base management theories: the CODASYL (network) approach; the hierarchical approach; and the relational approach. It was finally decided to employ the relational approach in the system's design because this model of data was seen to have a number of inherent advantages not shared by other models.

We shall not go into the theory of the relational approach here, but the bibliography cites several items on the subject for those who wish to know more about it.

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.

The Library serves the staff of the Centre and the staff of IDRC-funded projects in developing countries. It also serves interested public officials and researchers in Canada.

It has a data base of over 20,000 records representing both monographic holdings and analytics, and a serials data base of some 3,000 records. There are also some 4,000 records representing items on order or awaiting cataloguing.

Records are written in English, French or Spanish (the three working languages of the IDRC) according to the language of the document. Subject access to the collection is via on-line searching of descriptors assigned to each item from the IDRC's trilingual version of the Macrothesaurus (Organisation for Economic Co-operation and Development 1972).

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

MINISIS IN THE LIBRARY

To the user, the new system is, in some respects, similar to ISIS and in other respects different from ISIS - the latter is particularly noticeable in that MINISIS is designed for a multiuser environment. The three formerly separate data bases (acuisitions, serials, and monographs) are now merged to form a single data base. But user views have been defined so that each user sees the data base in the way most appropriate to him. Thus, the acquisitions librarians 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.

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 descriptors
embedded in an abstract. Once the cataloguing data has been proofread
and pronounced "clean", it is released by the data-base manager for
bibliographic searching.

There are also five automated authorities 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 code of institutions
(whether identified as a corporate author or an affiliation), that
correspond, for each bibliographic record, with the "place where the
work was 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 has been established
which eliminates "noise" words (prepositions, pronouns, conjunctions,
etc.) on request.

Now, let us look more closely at how the system appears to the
user. The system is fully interactive and a very important innovation
is the fact that aspects of computerised library operations which
formerly needed to be carried out by computer staff are now carried
out by library staff. 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 librarian creates his/her own 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. Because of the multiuser environment many users may use the same data base at the same time to do many different things.

The library user, having signed on to the system, and before choosing the portion of the data base on which he/she wishes to work, is asked to choose one of the six available functions (i.e. processors).

(1) ENTRY allows you to enter 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 desired field is 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. At present, vendor codes and institution-name codes are validated on line, while descriptors are validated in batch at the time the record is released for searching.

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 while the inputter is in the process of entering a new bibliographic record.

(2) MODIFY allows you to make 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. A query in MODIFY is a sub-set of the QUERY processor. 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.

(3)   INDEX allows you to sort records into various orders 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 IDRC Library), and can also handle the Spanish-language alphabetic sort sequence.

(4)   PRINT allows you to specify the format for printing each record (order and arrangement of fields; insertion of literals and punctuation) and to specify the page lay-out (number of records per page; whether printed in columns, tables, etc.). This function also initiates a print-out.

(5)   QUERY allows you to retrieve from a data base. Two kinds of search method are available - entering each operand together with its logical relationship to the other operands, or entering selected operands and combining them in a subsequent operation; the two methods can also be combined. Free-text and "inverted file" techniques with thesaurus expansion are also available during the retrieval process. Any validly defined field in the user's data base may be used for retrieval. The items found in a search can then be displayed on the screen and/or can be printed on the line-printer for delivery to the requestor. A set of instructions available to the user in QUERY allows one to save a search formula to be used again on a future occasion; save the records resulting from a search so that they can be used as a hit-file in the INDEX or PRINT functions; change to a different data base in the middle of a search; limit the search to only one or two of the three available languages; specify the print format desired for displaying and printing the records; and browse (display) the records.

(6)   RELEASE flags a record as available for searching once it has been declared "clean". The process can also be reversed to "unrelease" a record to modify it. No changes can be made to a "released" record.

A seventh function, which will be available at the end of September, is the arithmetic processor - COMPUTE. This will be used for such purposes as estimating the amount of money committed to books for a given part of a financial year, or for estimating financial commitments on IDRC-funded projects.

Also to be made available soon is an SDI processor.

Finally, the function EXIT is selected when you are ready to sign off the system.

The system handles all French and Spanish diacritics using the options available within the ISO 7-bit coded character set (International Organization for Standardization 1973); and it has a full upper and lower case capability.

The fact that all these functions are available to the user means that, now that the system is operational, only one half-time programmer will be needed to serve the library applications.

A set of utilities has been provided within MINISIS for the data base manager. For example, there are processors to produce and accept exchange tapes in ISO 2709 format (International Organization for Standardization 1973), to define users views for the system, to renumber records (useful when producing printed bibliographies), and for inverting on fields to which fast access is needed.

All code for the system (applications routines and data-base management routines) was written in SPL, the systems programming language of the HP 3000. The system runs on an HP 3000 with no special options.

It has been estimated that acquisition of equipment for such a system would cost between $150000 and $200000 - depending on amount of memory, number of disks and number of terminals acquired. The IDRC acquired 192 kilobytes of memory, one 1600 BPI tape drive, one printer with upper and lower case, one operator console and twelve terminals. Our investment represented the cost of running ISIS at the service bureau for two years. Of course, Hewlett-Packard has now brought out its improved Series III, on which MINISIS software will run just as well, if not better.

The design and implementation of the system took 45 person months, i.e. two persons for twelve months and three persons for an additional seven months. During the first twelve months, the design of the system was completed and thirteen of the seventeen data-base management routines were completed. During the next seven months, three of the four remaining data-base routines and the user processors were written.

In October, 1977, the first users - the acquisitions section - were phased in. By this time the data entry, modification and print processors were ready. In December, 1977, the whole data-base was successfully converted from ISIS to MINISIS. During early 1978, the query and index processors were completed, and the users' manuals written. At the time of writing, the arithmetic processor, the SDI processor and a photocompostion interface have yet to be completed.

The team's progress is proof that a small devoted development team is as effective, or even more effective, than a larger group.

Besides the library data base, the system supports two other bibliographic data bases - DEVSIS Canada (economic and social development information) and SALUS (information about low-cost rural health care and health manpower training). Development of a non-bibliographic application is also underway - a project information system (PINS) which will store and retrieve information concerning IDRC funded projects. PINS will be able to provide monthly expenditure reports to each division and to provide a question and answer service to management. Typical search questions would be:- How much money has the IDRC committed for projects on fertility regulation? How much financial support has the IDRC pledged to the International Rice Research Institute? Identify all recipient organisations in Latin America. All the processors available to library users will be available to the PINS manager. The arithmetic processor will play an important role here to compute the dollar figures.

The system also supports data-bases from other organisations. The ILO, FAO and UNIDO data-bases have been converted successfully to be searched on-line, while the Unesco data-base is expected soon. Until we can afford more mass storage, we handle the problem of on-line access through scheduling the non-IDRC data-bases for fixed times during the week or on advance user request.

It is hoped that MINISIS will prove to be attractive to institutions that require a bibliographic data-base system, but cannot afford expensive equipment and have no means of sharing a larger computer facility. Already, two large organisations - the International Labour Office in Geneva and the Agricultural University in Wageningen, the Netherlands - have expressed strong interest in MINISIS and the system was recently demonstrated to top-management and librarians in these institutions. MINISIS is a logical outgrowth of our experience with ISIS, and it already appears to be enlarging the common network that began with ISIS.

REFERENCES

CODASYL 1971. CODASYL data base task group report. New York, ACM, April 1971.

Codd, E.F. 1970. A relational model of data for large shared data banks. CACM, 13(6), 377-397.

1971. A data base sublanguage founded on the relational calculus.
In proceedings of 1971 ACM-SIGFIDEL workshop. New York, ACM.

1972. Relational completeness of data base sublanguages. In
Rustin, R. ed. Database systems, Englewood Cliffs, N.J.,
Prentice-Hall.

Date, C.J. 1975. An introduction to data base systems. Reading, Ma.,
Addison-Wesley.

Eswaran, K.P., Gray, J.N., Lorie, R.A., and Traiger, I.L. 1976. The
notions of consistency and predicate locks in a data base system.
CACM, 19(11), 624-633.

International Organization for Standardization. 1973. ISO-2709
Documentation - format for bibliographic information exchange on
magnetic tape.

International Organization for Standardization. 1973. ISO-646 7-bit
coded character set for information processing exchange.

Knuth, D.E. 1973. The art of computer programming, volume 3: sorting
and searching. Englewood Cliffs, N.J., Addison-Wesley.

Merrett, T.H. 1976. MRDS - an algebraic relational database system.
In Proceedings of the Canadian Computer Conference, Montreal, May
1976, 102-124.

Organization for Economic Co-operation and Development. 1972.
Macrothesaurus: a basic list of economic and social development terms.
OECD, Paris. (Also available in French, Spanish Portuguese, German,
Arabic and Indonesian.)

Ries, D.R. and Stonebraker, M. 1977. Effects of locking granularity
in a data base management system. ACM transactions on database
systems, 2(3), 233-246.

UNISIST/ICSU-AB Working Group on Bibliographic Description. 1974.
Reference manual for machine-readable bibliographic descriptions.
Unesco, Paris.