

64712

IDRC-Lib
64712

Information
Sciences
Archival Copy

MINISIS/UNIMARC Interface - Its Impact on Libraries

by

Dr. Charles A. Godfrey, Michael Sherwood and Garth Graham
(International Development Research Centre, Ottawa, Canada)

71415

Abstracts

MINISIS is a bibliographically-oriented information management and retrieval software package developed by IDRC for Third World institutions. It operates on Hewlett-Packard 3000 minicomputers and is currently used by 95 institutions in 22 countries. This paper describes the joint efforts of IDRC and IFLA to design and implement an interface to UNIMARC to allow for compatible exchange of information with libraries using other systems. The major technical problems encountered (fixed fields, repeatable subfields, etc.), the solutions chosen and implications for MINISIS users and other libraries are briefly reviewed. The interface, which will also offer some facilities for handling variants of MARC and CCF, will be ready for distribution to MINISIS users in 1985.

ARCHIV
GODFREY
NO. 5

MINISIS/UNIMARC Interface - Its Impact on Libraries

by

Dr. Charles A. Godfrey, Michael Sherwood and Garth Graham
(International Development Research Centre, Ottawa, Canada)

IDRC

The International Development Research Centre (IDRC) is a development aid agency created in 1970 and funded by the Canadian government, but overseen by an international Board of Governors. The principal mandate of IDRC is to support research into the application of science and technology for the socio-economic advancement of Third World countries. This goal is pursued primarily through project support to research and research-related institutions in developing countries. Project funding amounted to nearly \$80 million CAD in 1983. About 80% of IDRC projects involve research in the agricultural, health or social sciences with information science projects comprising a further 15%.

The Information Sciences projects most often assist in the establishment of new information activities in Third World countries to serve researchers and policy-makers. Less frequently projects involve information science research or the development or enhancement of information tools. The MINISIS computer software is the most notable example of information tool development by IDRC.

ISIS

The roots of MINISIS are found in ISIS (Integrated Set of Information Systems), a generalized information management and retrieval software

package developed in the late 1960's by the International Labour Office in Geneva. This software offered powerful capabilities for managing structured texts, including bibliographic records, on the IBM 360/370 series of computers. Demand by member countries for access to the software brought the package into the semi-public domain by the 1970's and developing country institutions approached ILO, UNESCO, IDRC and other agencies for assistance in the installation and implementation of the ISIS software, in most cases, for automation of specialized information services. ISIS is more efficient in the information retrieval, indexing and abstracting features demanded by special libraries than in the acquisitions, circulation and other library administrative functions needed by many libraries.

UNESCO has since assumed responsibility for ISIS and has developed and distributed widely its CDS(Computerized Documentation System)-ISIS version which operates on OS(Operating System)-type IBM system configurations. IDRC, on the other hand, developed and supported a DOS(Disc Operating System)-based version of ISIS until 1978 when its own operations were transferred to MINISIS. Although distribution has continued, a shared DOS version has not continued to develop as has CDS-ISIS.

MINISIS was developed by IDRC essentially to make the ISIS functions available in software which would operate on less expensive minicomputers and to incorporate into the software, features which would further meet the needs of an international and varied user community. Design and development of MINISIS began in 1975 and the first version of MINISIS (version A) was released in 1978. Currently, version F is distributed and the software is installed in more than 95 institutions in 22 countries. Commercial distributors supply direct services to developed country users and IDRC supplies the training and services to non-profit developing country institutions who receive the package at no charge.

MINISIS Design

MINISIS was designed to be functionally compatible with ISIS (Valantin, 1981), but its internal design is very different. Many factors were taken into consideration in the design work (Daneliuk, 1978). Among the foremost were requirements that the system possess:

- general applicability - the system should be as general purpose as possible;
- modularity - the system should be totally modular to promote ease of maintenance and extension;
- independence - the applications functions should be independent of the database management functions;
- 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;
- 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; and
- cost-effectiveness - the basic system must operate efficiently on minicomputers and be completed within 2 years (i.e. by 1978).

Following the extensive preparatory studies in 1975/1976, it was decided to implement the system on Hewlett-Packard (HP) 3000 minicomputers and to develop a relational model database management

package which satisfied the requirements. The software was written in SPL, the Systems Programming Language of the HP, and closely linked with the MPE (Multi-Programming Executive) operating system of the HP to achieve a high degree of efficiency in the software. As a result, the software is not portable to other types of minicomputers.

Currently, the MINISIS system consists of 20 processing modules which can be used online or in batch mode to manage and access databases. No programming is necessary to develop most user applications, though a few large users have done some limited programming. Databases are defined and generally managed using the DATADEF processor. The system is very flexible and database definitions can be extensively modified without reloading the data.

Some limitations are imposed by hardware and software on the complexity and size of records. No more than 256 unique fields and/or subfields can be defined for a database. A maximum of 9 subfields are allowed within a field and subfields are not repeatable within fields, though fields themselves are repeatable. It is these design limitations in particular that restrict the compatibility of MINISIS with full-scale MARC. Physical records have a maximum size of 4096 bytes, while logical records may be as large as 32Kb. Large files containing several million records are feasible although to our knowledge, the largest so far implemented in MINISIS is about half a million records.

End users access their databases using the processors ENTRY, MODIFY, QUERY, COMPUTE, INDEX and PRINT as required and permitted by the security profile established by the database manager. For example, new acquisitions records would be entered initially using ENTRY and any later updates (e.g. full cataloguing) would be made using the MODIFY processor. QUERY, which includes full Boolean operators, an SDI (Selective Dissemination of Information) subprocessor and many other facilities, might be used to select a subset of orders which can be sorted in INDEX and output to any device using formats defined with PRINT.

It is possible to manage most library functions using a sequence of the end-user processors, but some operations such as circulation cannot be managed efficiently. Therefore, a few large MINISIS installations have developed their own specialized processors to meet their specific needs. Examples are serials check-in, public catalogue access and circulation processors. When these user-made programs are sufficiently developed, they become available to other MINISIS installations through a user-contributed library of software.

Just as its predecessor ISIS, MINISIS is most suitable for libraries of small to medium size with a greater need for powerful retrieval capabilities than for intensive administrative functions.

Information Exchange

Few bibliographic computer systems use the same record format for internal record storage and processing as is used for the exchange of records on magnetic media. A conversion is done when data is written to output devices for exchange with another system. This is also true of MINISIS, though it is somewhat unique among database management systems in the use of a bibliographic-type internal format. The internal record format of MINISIS was designed to conform generally with the ISO 2709 (International Organization for Standardization format for bibliographic interchange on magnetic tape) record structure. The tag numbering (letter plus 3 digits) system and selected other elements of the record structure were adopted from the UNISIST Reference Manual (Dierickx and Hopkinson, 1981). MINISIS handles data in extended and non-Roman character sets up to 256 characters that conform to 7- or 8-bit ASCII (American Standard Code for Information Interchange) code.

A specific processor named ISOCNV was developed to import and export bibliographic data to and from other systems, in batch. This module is driven by user-defined correspondence definitions and was generalized to accept and generate records in a variety of ISO 2709 based formats including ISIS, AGRIS, MARC's and others. Initially, however, only conversions to and from ISIS formats were fully implemented in the processor. Later, a further processor called BATCHIN was added. This processor accepts ASCII files in the format "tag: data, tag: data" and is a convenient means of exchanging all types of data.

Much experience has now been acquired by MINISIS installations throughout the world with a variety of bibliographic data exchange formats. In most cases, MINISIS has proved to be quite hospitable to external data and very capable of outputting data in formats which can be processed by other systems. However, it has only been partially successful in handling MARC data as first demonstrated by the experiences of the Library of the National University of Singapore (Hochstadt, Quah and Ong, 1982) as well as the experiences of Systemhouse, the North American distributor, and some of its sublicensees. Importing data from MALMARC (Malaysian MARC) tapes proved easier, though not fully satisfactory, than exporting data to that format. This was recognized as a serious limitation to the system for many libraries in developing countries.

UNIMARC Interface

In 1982, the International Federation of Library Associations and Institutions (IFLA) opened discussions with IDRC to explore ways in which both organizations might cooperate to enhance MINISIS to handle input from and output to UNIMARC format (Hopkinson, 1983). Coordinated by Henriette Avram on behalf of the IFLA Program Management Committee and with support from the British Library,

Library of Congress (LC) and the National Library of Canada, a consultant, Elaine Woods, was contracted in 1983 to conduct a study on the feasibility of developing a UNIMARC interface to the MINISIS package (Woods, 1983). At the suggestion of IDRC, the new and untried Common Communications Format (CCF) (Simmons and Hopkinson, 1984) which was being developed by UNESCO, was also included in the study.

The principal objectives of the study were to determine to what degree MINISIS is capable of implementing UNIMARC/CCF and to specify which changes could be made to MINISIS to achieve maximum implementation of UNIMARC. The study is detailed and thorough and provides an excellent basis for implementation.

Briefly, the study revealed that, due to original design decisions, it is not possible to efficiently implement full MARC using MINISIS, but the most frequently used MARC fields and features can be implemented with certain modifications and enhancements to MINISIS. The principal problem areas in MINISIS turned out to be:

- (1) the limit on the number of tags;
- (2) the non-repeating and limited number of subfields; and
- (3) limited capabilities for handling fixed fields.

Interface Design

A second project with IFLA is now underway to implement a UNIMARC interface in MINISIS by the end of 1984. This work is being carried out by IDRC computer staff in collaboration with Elaine Woods whose

participation is being generously supported by the Council on Library Resources (CLR).

The primary objective is to develop a UNIMARC interface which will make MINISIS more effective for developing country libraries and will encourage the MARC libraries to adopt UNIMARC as their common exchange format. However, IDRC has decided to generalize the interface to possibly handle other variants of the MARC format. Just as developed countries, many developing countries have adopted national or regional MARC-based exchange formats. Though MINISIS will not promise a compatible interface to each of these, it is expected that the interface will accommodate some with no additional programming by the library and others with a minimum of programming.

Several modifications and enhancements are being made to MINISIS to create the interface to UNIMARC. Firstly, the ISOCONV processor will be extended to handle additional leader elements used in many MARC formats such as record status, type of record, encoding level, etc.

Secondly, ISOCONV will also be extended to handle a special archival field. This will be a user-defined repeatable field in MINISIS that stores, in original form, the tag, indicators and data of fields received in external records which are not needed by the library and are not uniquely defined in their MINISIS database. Alternatively, data required only by an external system could be input directly into the archival field in a MINISIS record for later output to an exchange tape.

Thirdly, table-driven procedures will be added to handle fixed fields with far greater flexibility and efficiency than is currently possible in MINISIS. The user will be able to define and manage multiple (not repeatable) fixed fields such as the UNIMARC fields 100, 105, 110, 115 and 12x or their equivalents in other MARC formats.

Fourthly, special facilities will be added to several MINISIS processors to handle a limited number of MINISIS fields with non-MINISIS subfield delimiters embedded in them. The processors (and especially the display intrinsic) will be able to call a table-driven routine for each field that requires special manipulation of the delimiters and data. Repeatable subfields and more than 9 subfields will be allowed in a field. The table will offer a finite set of basic options for each subfield to generate keys, sort or replace delimiters with user-defined strings (punctuation) each time the record is printed.

Many MARC fields can have repeatable or more than 9 subfields, though the frequency that this occurs may be very low in practice. Because this mode of operation will add a significant processing overhead to MINISIS, this technique should be restricted by the user to the more significant or frequent UNIMARC fields (200-title statement, 500-uniform title, 600-607 subject fields, 670-PRECIS).

MINISIS is not efficient in this specific situation because it has been optimized for general applicability rather than to meet the requirements of a specific application or format. It is also for this reason that MINISIS cannot match the specialized library software packages in many library applications.

It has also been necessary to leave a few UNIMARC components out of the interface. Indicators will not be implemented, though users will have the option of storing indicators in a MINISIS subfield if there is sufficient space in the data definition. Also series/linking fields will not be accommodated, though they can be carried in the archival field. Nor will the system manage the UNIMARC links between records containing the analytic and generic document, though in some cases, this may be possible using the relational joins in MINISIS.

(

User package

About 30% of the enhancements had been programmed as of July, 1984. When the remainder have been completed and tested in-house, they will be field-tested in a pilot project.

In 1985, the UNIMARC software enhancements will be incorporated into the next standard release version (G) and MINISIS user documentation which are distributed to all installations. Most of the special routines will be supplied to users in source code to enable them to make any further modifications or enhancements needed to meet their specific requirements. In addition, a special manual will be prepared describing the MINISIS correspondence and data definitions and table values which will yield the highest compatibility with UNIMARC. As always with MINISIS, the actual implementation remains the responsibility of the user library which possesses the full flexibility of MINISIS to meet their own needs.

Implications

Libraries that use the MINISIS/UNIMARC interface fully will acquire several capabilities that do not already exist in MINISIS. They include the following:

1. data from UNIMARC records (and perhaps other types of MARC records) can be converted to a MINISIS database without any loss of data and without any programming;
2. fixed fields can be managed in an efficient and user-friendly fashion;
3. the title and statement of responsibility can be compatible in great detail with the conventions of the Anglo-American Cataloguing Rules, Second edition (AACR2) and the International Standard Bibliographic Description (ISBD);

4. subject headings, for example, Library of Congress (LC) headings, can be fully implemented;
5. locally captured data can be converted to UNIMARC records and exchanged with other libraries; and
6. new options in the system can be used to manipulate local fields as well.

Though MINISIS will still not offer the same range of cataloguing facilities that are available in some other bibliographic packages, the UNIMARC interface does represent a major improvement for libraries and is essentially as far as MINISIS can specialize in this direction. The cooperation offered by IFLA to IDRC has ensured that developing country libraries who use MINISIS, perhaps for its ability to handle multiple and special character sets or perhaps for its other features, will not be impeded from exchanging information with libraries world-wide that accept UNIMARC records.

REFERENCES

- Daneliuk, F.A.; Information Retrieval and Library Management: an Interactive Minicomputer System. Ottawa, Ont., IDRC, 1978. 16 p.:ill (IDRC-TS14e). ISBN 0-88936-170-3.
- Dierickx, H. and A. Hopkinson (eds); Reference Manual for Machine-Readable Bibliographic Descriptions. 2nd rev. ed. Paris, UNESCO, 1981. 341 p. (PGI-81/WS; 22).
- Hochstadt, P.W.C., Quat, J. and G.H. Ong.; Towards an Online Integrated System at the National University of Singapore Library. Presented at 41st Congress, FID, Hong Kong, 12-16 September 1982. 24 pp.
- Hopkinson, A. (ed); UNIMARC Handbook. London, IFLA, 1983. 264 p. ISBN 0-903043-40-8.
- Simmons, P. and A. Hopkinson (eds); CCF: The Common Communication Format. Paris, UNESCO, 1984. 185 p. (PGI-84/WS/4).
- Valantin, R.L.; CDS/ISIS and MINISIS: a Functional Analysis and Comparison. Ottawa, Ont., IDRC, 1981. 88 p. (IDRC-TS37e). ISBN 0-88936-296-3.
- Woods, E.W.; Report on MINISIS/UNIMARC study. Unpublished, Nov. 1983. (Available on microfiche from IDRC).