CARIBBEAN DEVELOPMENT AND CO-OPERATION COMMITTEE

THE CARIBBEAN REGION

ECONOMIC COMMISSION FOR LATIN AMERICA AND THE CARIBBEAN
Subregional Headquarters for the Caribbean

CARIBBEAN DEVELOPMENT AND CO-OPERATION COMMITTEE

UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND
    CULTURAL ORGANIZATION

CARSTIN Training Workshop/Seminar on
Network Development in the Caribbean
Port-of-Spain, Trinidad
3-14 December 1984

COMPUTERIZED BIBLIOGRAPHIC

INFORMATION SYSTEMS

Prepared by

Fay Durrant
Consultant

UNITED NATIONS

ECONOMIC COMMISSION FOR LATIN AMERICA Office for the Caribbean

COMPUTERIZED BIBLIOGRAPHIC INFORMATION SYSTEMS

As computerized systems are being increasingly employed in the development of information services, and as such services range from single libraries, co-operative networks of libraries, national information systems, bibliographic utilities, and transnational suppliers or providers of databases, it seems useful to examine the basic objectives and functions of these systems, their current and possible effects on the library and information services, and their potential for the development of specialized information services in the Caribbean.

This paper will therefore examine briefly, the various types of computerized systems and networks, which:

- collect data;
- store and process by automated methods; and,
- retrieve and disseminate information.

The main consideration will be bibliographic information systems, but as consideration of the access to, and use of bibliographic and non-bibliographic information is becoming so closely related to that of statistical, numerical and full text, it will also be necessary to include here, consideration of non-bibliographic systems.

The information systems under discussion here perform the following activities:

- acquisition of primary documents (books, journal articles, patents, dissertations, technical reports);
- preparation of corresponding bibliographic records;
- determination of appropriate retrieval keys; and,
- dissemination of information (usually by a combination of printed products and/or computer search on request).

The various units performing these functions could be considered to be sub-systems which need to be interconnected to provide a nation or region with an integrated information service. The location of the national boundaries indicate that the design of the systems, which have national, regional, or international impact will be affected by political, social and legal situations and will need to take certain factors into consideration.

The technical considerations of information transfer and the corresponding requirements for national policy need to be determined, along with the national and regional problems of legal, political and language boundaries which need to be overcome. The International Institute of Communications in its assessment of the Use of Satellite Communication for Information Transfer which was prepared for UNESCO in 1982, identified relevant policy issues relevant to information transfer as:

- existence of an institutional framework, and mechanisms for the elaboration of national strategies in telecommunications and computerization;
- manpower development, continuing education and research;
- assessment and acquisition of technology and a hardware procurement policy;
- analysis of users' needs;
- planning of telecommunications systems;
- planning of computerized information systems;
- privacy in relation to individual or corporate data;
- standardization;
- international co-operation; and,
- trans-border data flows.

The harmonization of these policy issues on a national and regional level, would therefore encourage the integration of the various components of the complete information transfer cycle, and could lead to the development of a regional network such as EURONET.

In the absence of a national or regional policy, or while one is being developed, the issues of:

- standardization of methods of processing; and,
- standardization of computer hardware and software;

need to be taken into consideration in the design of any computerized network, or in the employment of the facilities of any commercially available system as the existence of compatibility - standardized or functional - will enable each system or sub-system to optimize existing facilities or opportunities.

An example of opportunities which are close to home, is the possibility of access to the CARISPLAN database. An examination of the processes of CARISPLAN shows that the database has been prepared on an IMB 360 at ECLAC, Santiago in IBM format, but has been translated into a suitable format for the Radio Shack on which it now runs at ECLAC, Port-of-Spain. There is no standardization between the IBM machine and the Radio Shack and the degree of functional compatibility between the two systems is therefore related to the complexity of the program which has to be written for translation of the records from the IBM format to the Radio Shack. National focal points and participating centres with similar or compatible hardware and software currently in use at ECLAC would be able to obtain copies of the CARISPLAN database and to use them for providing services to their own institutions.

The advantages of an automated rather than a manual system are undoubtedly evident to all experienced librarians and I think that the concern of professionals in the region is rather the selection of appropriate computerized systems which accomplish the most efficient processing, while ensuring compatibility with other systems in the region and permitting access to internally available sources of information.

The developments in computerized bibliographic information systems have resulted in significantly more flexible user friendly systems, employing less rigidly structured programs and able to run on smaller more compact machines. Input procedures have been greatly simplified permitting more efficient data entry, without reduction of the retrieval capabilities. The databases of these systems have therefore been able to provide more flexibility in searching, and consequently more sophisticated products mainly in the form of Current Awareness Services, and Selective Dissemination of Information. The telecommunications networks which have now been expanding their facilities and the development in communication, now permit access to software by personal computers to the on-line large range of bibliographic and numerical databases.

These development seem to indicate much greater likelihood of individual libraries, and networks in the region having access to computers and being able to install systems and/or being able to expand their services by linking networks with external sources of information.

POTENTIAL FOR THE CARIBBEAN

The introduction of computerization into the region, although already initiated is still in its infancy in terms of bibliographic information systems. It therefore seems to be a most appropriate time to examine the systems which are to be installed in terms of established criteria and in terms of taking advantage of the best aspects of the state-or-the-art.

The information technology environment here is certainly becoming more and more influential, with the development of automated techniques replacing manual practices in offices and other commercial enterprises. We will discuss later the possibility of transposing the office filing system programs into use for library databases, which are less well developed.

The development of national information networks contains two main areas in which computerization would significantly enhance retrieval and improve the quality of the services provided.

The establishment of a computerized directory of libraries and information units in the country would provide users with access to current information on the subject areas covered by the libraries within the national system.

Several countries have compiled such printed directories manually, but it is virtually impossible to keep this perpetually up-to-date and therefore provide current information for retrieval.

There are several simple filing and data base management packages which would be able to accommodate this information, allowing the data to be stored on microcomputers, regularly updated and searched as required and even distributed to participating libraries. An extract of the Jamaica directory will be demonstrated at the CDC.

The second area which would be an extension of the referral service, is the automated equivalent of the national union catalogue. The determination of a suitable system - software and hardware - to be phased in, and the preparation in standard format of machine readable records, by all units within the national system would be the main step required for an automated version of a national union catalogue. Several countries are now in the process of determining an appropriate system for processing material, and for eventually providing the facility of a union catalogue which in its manual form is too unwieldy to contemplate.

ESTABLISHING SYSTEMS

The impetus towards computerizing information systems in the region is likely to come from various angles. Initially, the library of information unit's parent organization is likely to encourage use of computers - often minis, micros which are being purchased or acquired as gifts by the organization.

In terms of large organizations, the Caribbean Development Bank's Computer Centre has, since 1981, installed an IBM 4331 mainframe machine which runs CDS/ISIS - Integrated Set of Information Systems - a package of programs for information processing. This is the only installation of this scale which is expected in operation in the region, but it is expected the National Library of Jamaica which is contemplating the installation of another version of ISIS at that country's National Computer Centre, will have a similar system in place in 1985.

The existence of these two large-scale computerized systems will create an additional impetus, while possibly establishing de facto standards for future systems.

In the smaller organizations, micro and mini computers have begun to appear. These are often available to the library or information centre. Again the main reason for their not being immediately employed in some aspects of information retrieval, is the lack of appropriate software, which can be employed, and in some cases the need for computer literacy of professionals, and/or need for trained technicians and data entry personnel.

Some of the impetus for installation of computerized bibliographic information systems has also come from the hardware houses, which have intensified their marketing strategies mainly in the area of office automation - management information systems, dedicated word processing and accounting information systems.

The Caribbean Information System for Economic and Social Planning - CARISPLAN is an example of a decentralized computerized bibliographic information system. The main processing is done by CDS/ISIS on an IBM 360. As the participating centres do data entry on printed bibliographic record cards, this has provided a format for their in-house processing in antici-pation of automation of their libraries.

Already a fairly wide range of makes and models immediately has implications for compatibility. The fact that the prices of hardware have also been reduced is also likely to lead to more purchases at least of micros, some of which are available for less than US$1,000.

In consideration of the establishment of a regional bibliographic network, and corresponding use of computers, this obviously raises the question of the feasibility of linking the sub-systems, of library and information units or the possibility of employing similar software, and of transferring copies of databases from one library to the next.

In most cases, transfer of databases on diskettes requires indentical hardware, software, and diskette formats, or conversion programs for modifying the structure of the data.

The development of a regional network would require, at least in the future, telecommunications links between participating countries, to facilitate the flow of data. Although some of the national telecommunications facilities now include packet switching, this is only gradually being introduced in the region, and to date the situation is:

COUNTRY	PACKET SWITCHING INTRODUCED/PROPOSED
Antigua	1985
Barbados	1982/3
Jamaica	1985
Trinidad	1984

which signifies that there could not yet be a genuine network established from the pattern of EURONET, but rather a network based on a combination of techniques, possibly computerization along standard lines at the local level, and with subsequent linking into an on-line network when the telecommunications facilities have been expanded to permit appropriate communication between member countries.

The telecommunications networks TYMNET, TELENET, UNINET, - being the main ones, which can be accessed through the external communications agencies, are now being used as a more efficient means of searching overseas databases. These databases are now supplied through vendors, hosts or database suppliers and now provide:

- bibliographic references;
- abstracts of entries;
- full texts of entries; and,
- numerical and statistical data

on a range of subjects.

PLANNING FOR COMPUTERIZATION

The determination of an appropriate system involves an initial process of systems analysis, determination of the requirements for the system. The first step would be the selection of software - likely to be able to provide the requirements. In the design of modern bibliographic information systems, analysts and librarians tend to aim at the use of integrated software packages - applications programs with which untrained users can interact easily. This generally implies the use of a high-level programming language Such as FORTRAN, COBOL, ALCOL, BASIC or PASCAL.

The choice initially is between a set of user-written programs, which cover the various applications, and a ready-made package which have been prepared by an international agency or commercial firm and have been preferably tested according to certain standards and criteria. The wide availability of the ready-made packages aids in the evaluation which would be necessary, as the journal literature is likely to have examined the program in operation, and to be able to then assess their operating capacity and their effectiveness in comparison with other available systems. Ashford, for example, in an article in Program, April 1984, examined several software packages which are in use in libraries and information units, and assesses their capabilities. Gerald Lundeen and Carol Tenopir do a similar analysis for microcomputer-based software packages in Microcomputers for Information Management, September 1984.

Much of the information about such packages is unfortunately available from the manufacturers' brochures and announcements which might not be unbiased, and might not reflect testing in a real library situation. Reports of innovations proposed and claims of new capability which have been introduced need to be very carefully assessed, as the balance between a well-tested and an innovative system certainly needs to be achieved.

The structure of the databases created by these packages, needs to be considered in order to select the package most suitable to the needs identified earlier.

As the databases created in bibliographic information systems, can be roughly classified into free-text or unstructured and structured, an examination of the two groups and comparison with the needs and priorities of the proposed system, will determine the type of database structure which is most suitable.

Bordwell in his recent article on dBase II in Program, examines the possibility of using dBase II for retrieval in a bibliographic information system, and concludes that 'commercial' database management systems are of limited applicability in the establishment of bibliographic databases. He justifiably illustrates this by comparing the narrow combinations permitted by a basic DBMS structure with the flexible retrieval search strategy of on-line bibliographic databases.

The flexibility in searching might range from requiring a single search key to permitting a fairly narrow combination of terms.

In comparison with the major commercial on-line databases such a search strategy would seem primitive and in fact a DBMS employed for the retrieval of bibliographic information could only be employed regularly with the use of a query language.

The MINISIS software developed and maintained by IDRC provides an example of a database management system in use for information retrieval and library management. In recognition of the fact that relational theory does not provide for non-relational access paths to a database, a B-trees storage technique was used to implement inverted files, while compacted bit maps were used to store postings and to implement the Boolean logic.

The free-text or information retrieval packages used for information retrieval also show the characteristics which have been specially programmed into the MINISIS system. These packages, a number of which were analysed and evaluated by Ashford in his article on text-based retrieval packages, exhibit certain common characteristics which also make for greater flexibility in searching and for quicker retrieval of information.

Of the nine systems reviewed by Ashford, all provided the following retrieval functions:

- Boolean query formulation (AND, OR, NOT);
- search refinement (by sub-questions);
- field or section limits on serach scope;
- field or chapter limits on serach scope;
- display or 'postings' frequencies;
- truncation of word termination;
- text string searching;
- stored searches or stored questions;
- thesauri including acceptable and unacceptable terms;
- synonym recognition;
- help for users;
- sorting of search results;
- editing of search results;
- formatting of output;
- security controls at database.

These retrieval capabilities reflect the file structure, which would be required, and also indicate a useful 'shopping list' for softwear selection.

Selection is also likely to be influenced by:
- demonstration in a working environment;
- vendor support (proven and potential for the future);
- consistency in marketing policies;
- availability of hardware; and,
- ability to accommodate users requirements.

The main international standard which would be related to the use of software within networks is the ISO 2709 for exchange of bibliographic data. The advent of a standard in relation to microcomputers or in the modification of the 2709.