

09436.00

(21142)

~~CE~~
E

Va al Riel

*** DRAFT 12/01/83 ***

A GUIDE TO THE
COMPUTERIZATION OF POPULATION DOCUMENTATION
STORAGE AND RETRIEVAL SYSTEMS

Written at the request of the
Population Information Network
(POPIN)

Arthur M. Conning
CELADE
Casilla 91
Santiago, Chile

(File: POPIN4.PUB)

EXPLANATORY NOTE

This draft of the Guide has only two sections:

- (1). An executive summary, and,
- (2). A review of options, technical problems, etc., for the Population Information Center.

It does not follow the original terms of reference, which include a third section geared to telling the information worker "how to carry out the operations described".

It soon became clear that a Guide on computerization is very different from one on, for example, information acquisition, thesaurus use or classification schemes. Computerization requires that the manager and his or her staff learn a whole new vocabulary and philosophy on the storage and retrieval of bibliographic information. Consequently, a rather lengthy section was needed just to present the basic concepts and to provide the information that a manager needs to decide whether to computerize, how to make key decisions on the design of the system and how to analyze alternative options.

Fortunately, the extra length of the managerial section could be compensated by eliminating the "how to do it" section, since the latter would have been useless if it did not treat a large number of different topics in detail, which was obviously impossible in anything less than a full-length manual. Various references are given to meet this need.

TABLE OF CONTENTS

	Page
Executive Summary	iii
1. INTRODUCTION	1
2. BASIC COMPUTER CONCEPTS	2
2.A Computer components p. 2	
2.A.1 Hardware p. 2	
2.A.2 Software p. 4	
2.B Communication with the computer p. 4	
2.C Categories of computer p. 5	
3. COMPUTER STORAGE AND RETRIEVAL OF BIBLIOGRAPHIC INFORMATION	5
3.A Creation of secondary bibliographic information p. 6	
3.B Computer Retrieval of bibliographic information p. 6	
4. ADVANTAGES AND DISADVANTAGES OF COMPUTERIZATION	7
5. COMPUTERIZATION OPTIONS	8
6. CRITERIA FOR EVALUATING THE CONDITIONS AND DESIGN REQUIREMENTS FOR COMPUTERIZATION	9
6.A How many documents? p. 9	
6.B Who are the users? p. 9	
6.C How many users are there and where are they? p. 10	
6.D What services are required? p. 10	
6.E What "computer environment" is required? p. 10	
6.F What resources are required? p. 11	
7. BASIC DECISIONS IN THE DESIGN OF A COMPUTERIZED SYSTEM	12
7.A Flow diagrams of Input and Output operations p. 12	
7.B Decisions concerning Input p. 12	
7.B.1 Scope definition p. 12	
7.B.2 Bibliographic levels and documentary units p. 13	
7.B.3 Bibliographic record content p. 14	
7.B.4 Bibliographic description p. 14	
7.B.5 Selection of international and internal formats for bibliographic description p. 14	
7.B.6 The internal bibliographic record format and the PI Center worksheets p. 15	
7.B.7 Abstracting and indexing p. 16	
7.B.8 Data entry, correction and modification p. 17	
7.B.9 Inverted file p. 17	
7.B.10 Database security p. 17	
7.C Decisions concerning Output p. 18	
7.C.1 Searches p. 18	
7.C.2 The production of journals and listings p. 18	

7.D Decisions concerning Hardware and Software	p. 18
7.D.1 Computer/software combinations available	p. 19
7.D.2 The micro-computer/software solution	p. 20
7.D.3 Peripheral devices	p. 21
8. UTILIZATION OF A MANUAL SYSTEM TO FACILITATE FUTURE COMPUTERIZATION	22
9. UTILIZATION OF EXTERNAL COMPUTERIZED POPULATION DATABASES	23
BIBLIOGRAPHY	25
TABLES AND FIGURES	27
Table 1. 1A: Development and database creation costs. 1B: Operational costs.	
Table 2. Important computer/software (non-commercial) combinations for bibliographic information storage and retrieval.	
Figure 1. Simplified Diagram of a computer.	
Figure 2. Telecommunication links with a computer.	
Figure 3. Simplified diagram of a computer record and its fields.	
Figure 4. Serial and inverted file searches.	
ANNEX 1: Input and output flow diagrams	32
Annex 1A. Input flow diagram: Selection of documents.	
Annex 1B. Input flow diagram: Processing of documentary units.	
Annex 1C. Output flow diagram: Interactive searches.	
Annex 1D. Output flow diagram: Production of a journal.	
ANNEX 2: Search logic	35
ANNEX 3: Worksheets	38
Annex 3A. Worksheet used by CEPAL and CELADE/DOCPAL.	
Annex 3B. Worksheet recommended by IDRC.	
Annex 3C. Manual system workcard and UNITERM card compatible with the CELADE/DOCPAL computerized system.	
Annex 3D. Manual system worksheet used by RIPS/PIDSA.	
ANNEX 4: External search services	45
Annex 4A. POPLINE Literature Search Request Form and an Example of a Search Listing.	
Annex 4B. CELADE/DOCPAL Literature Search Request and Document Copy Request Forms.	
INDEX: Pages on which terms are defined or utilized substantively	50

EXECUTIVE SUMMARY

The computerization of the storage and retrieval of bibliographic data in a Population Information (PI) Center requires that the manager of the Center first learn a new approach to information management and the provision of population information services and then make a series of key decisions. Consequently, the objectives of this Guide are to acquaint managers and their upper-level documentalist staff, particularly from small PI Centers in developing countries, with the new concepts and vocabulary required and to assist the manager to make the decisions on whether and how to computerize. This Guide should be read with the understanding that the provision of the best possible service to users, consistent with their needs and the resources available to the PI Center, is not necessarily a commitment to use present computer technology.

Computers and the storage and retrieval of information

The equipment ("hardware") comprising a mainframe, mini- or micro-computer require instructions (computer programs or "software") on what to do. Using interactive software, an on-line bibliographic search allows the user to exchange questions and answers with the computer from a distant terminal.

In a computerized system, the description of a book, chapter, article or other document is written on a computer "master record", which is roughly equivalent to a card in a traditional library catalog. All the master and other associated records are called a "bibliographic database". With appropriate technology (an inverted file), a search of the entire database can be made very rapidly to find only the records meeting a user's request.

Advantages and disadvantages of computerization

The major advantages to the user of services from a computerized database are the vastly increased possibilities to locate specific information by combining any of the data elements on the master records and the receipt of printed results. Unlike a card catalog, a database can be the source of new services at little extra cost. The increased ability for information exchange with other centers also augments the user's access to population materials.

Computerization in many small PI Centers, however, will have various disadvantages. Since access to bibliographic information will be dependent on the computer, unreliable operation due to electricity shortages and other causes can lead to user and staff frustration and reduce the overall quality of service. The PI Center will also be dependent on programmers to modify software functions. Finally, development costs are likely to be high and the cost of operation is not likely to be less than previously.

Computerization options and decision criteria

Four major options should be considered:

- 1). Computerize as soon as possible;
- 2). Continue with the manual system because it is adequate and computerization is not justifiable in terms of volume or need;
- 3). Change to a modern manual system oriented to future computerization;
- 4). Use the computerized facilities of external database services.

The following questions will assist managers to decide on which option to select and on the design of a computerized system based on the needs and characteristics of the PI Center's users and the resources available:

- (a). How many documents will enter and what is the rate of increase?
- (b). Who will be the users?
- (c). How many users will there be and where are they?
- (d). What services are required?
- (e). What "computer environment" is required?
- (f). What human and financial resources are required?

Recommendations on the design of the input and output operations

- (a). Obtain expert advice on the system design.
- (b). Create a separate record for each part (e.g., chapter) of a document within the database scope to permit the location of specific information.
- (c). The description of the characteristics and content of a document and its organization on the computer record should be compatible with the format of the Regional Population Information Center, which in turn should be compatible with the UNESCO exchange format in the UNISIST Reference Manual (UNESCO RM, 1981).
- (d). The worksheets, used to fill in the bibliographic description (title, author, etc.), an abstract and keywords taken from the Population Multilingual Thesaurus (CICRED, 1979), should be essentially the same as those of the Regional Center, making local adaptations as necessary. If a worksheet is not available, obtain one from a POPIN associated regional center elsewhere.
- (e). After entering the data into the computer, it should be organized for efficient searching and proper security measures should be taken to avoid accidental database loss.
- (f). The search software should be interactive and "user-friendly".
- (g). An abstract journal should follow the format of the Regional Center journal, if it has one, to save development costs.

Recommendations concerning hardware and software

- (a). Obtain expert advice on hardware and software matters.
- (b). Storage and retrieval software should be obtained from elsewhere. Unless a PI Center has extensive resources and ample experience with databases and computer programming it should not attempt to write its own software.
- (c). The preferred computer/software combinations at present are an IBM main-frame computer with CDS/ISIS from UNESCO, Paris and the HP Series 3000 mini-computer with MINISIS from IDRC, Ottawa; these agencies will provide back-up support to user centers.

(d). If these systems are too costly and neither an IBM nor a HP 3000 is available, most PI Centers should postpone computerization until software for low-cost micro-computers can be obtained or until an adequate local agency can provide reliable computerized input and output services for the PI Center a reasonable cost.

Other options

(a). If computerization is not justifiable or is not possible, but improved information services are required, manual procedures using modern retrieval techniques, and compatible with a computerized system, should be introduced. The approach should be that recommended by the Regional Population Information Center. An example using a workcard and UNITERM card is given.

(b). All PI Centers, whether computerized or not, should take advantage of external computerized population database services. At present there are three such services with international coverage: (a). POPLINE, with over 100,000 records and abstracts in English, provides free searches to developing countries; POPULATION BIBLIOGRAPHY, with over 47,000 records and abstracts in English, is available from the commercial on-line search service DIALOGUE; and CELADE/DOCPAL, with around 17,500 and abstracts in Spanish on Latin American and Caribbean documents, provides free searches and a document copy service.

1. INTRODUCTION

This Guide is designed to assist managers of Population Information (PI) Centers - which may be libraries or documentation centers within larger parent agencies - to make decisions on whether and how to computerize their documentation operations. At the same time it should serve to acquaint managers and their staff with the concepts and vocabulary used when discussing and planning computerized documentation systems.

It is assumed that the PI Center is dedicated to the provision of the best service to its users, taking into account their needs and the financial and human resources available to the PI Center. The Guide does not presuppose that computerization is the correct approach for all PI Centers considering it. On the contrary, it is recognized that the commitment to provide the best service possible with the resources available is not necessarily a commitment to use computer technology since that will depend on the local environment in which the PI Center must operate and the specific services that its users require (Habitat, 1975:75).

There are a variety of tasks that can be partially or fully computerized in a library or documentation system:

Acquisition	Mailing and distribution
Serials control	Education and training
Cataloging	Administration
Circulation	Data processing and calculations
Inter-library loans	Storage and retrieval of
Word processing	bibliographic information
Translation assistance	

Of these only the last two can be considered substantive applications of the computer to population information. Furthermore, with the exception of word processing, which will be useful to almost anyone having to write and edit letters and documents, computerization of the other areas will be unnecessary or of very low priority for most small specialized population libraries and documentation centers (recent comprehensive guides covering these other areas are Boss, 1979 and Corbin, 1981, which are very useful handbooks although oriented to computerization in large U.S., Canadian and European public and university libraries).

The Guide has a number of deliberate limitations. First, the Guide will focus only on the computerization of substantive aspects of population documentation, in particular, on the computerized storage and retrieval of population information to provide specialized bibliographies on request from individual users and/or the generation of journals and lists. Second, since the problems involved in computerization of population information services will depend on the size of the PI Center and the resources available, and this short guide cannot treat all situations, it is primarily directed toward the problems of small, specialized, relatively low-budget population information units in agencies in developing countries.

It is important to be very clear on the type of population information that is being considered here - bibliographic material - since computer manipulation of this form of information is different from other frequently utilized forms, tabular data and micro-data. These three types, and how they are manipulated by the computer, may be distinguished in the following manner:

Bibliographic material ^{1/} includes books, chapters, reports, journal articles, conference papers, and other "printed" documents, whether published or unpublished. Audio-visual materials also can be considered to be a type of document. When a search for bibliographic materials is carried out, the individual document is the unit of retrieval, i.e., the unit of interest to the user.

Tabular data or statistical data include such items as computer-produced tabulations of census, survey and other micro-data (see below), time series, tables and graphs. The data in a population table, such as the numbers of migrants to the capital city by sex, age, and education, is normally about categories of persons or objects, rather than about any individual person or object. When the table is the unit of retrieval in a collection of tables, then tabular data may be treated as a special type of document, although unlike a document, after locating a table it may be manipulated by computer for statistical analysis.

Micro-data refers to the information (e.g., age, sex, and number of children ever born alive) obtained from a person in a census, survey, etc. All the data on all the persons in, say, a survey, constitute a micro-dataset. Since the individual case is not of interest in micro-data, information in a micro-dataset is aggregated by computer programs to obtain tables or other statistical output.

The guide will be concerned only with the use of the computer to retrieve documents. The methodology and computer programs for the analysis of tables or for the aggregate retrieval of micro-data, which are very different from those for bibliographic retrieval, are not treated here (a general outline of survey data processing is given in United Nations, 1982).

2. BASIC COMPUTER CONCEPTS

2.A COMPUTER COMPONENTS

2.A.1 Hardware

A computer is an electronic device that can be instructed to perform organized sequences of operations on data at very high speeds following instructions provided by the user. The physical parts of a computer are called computer hardware. Irrespective of its size, a general purpose digital

^{1/} An index is provided at the end of the Guide to facilitate looking up the definitions and utilization of the terms underlined in the text.

computer, which is the type of interest to PI Centers and their parent agencies, is normally composed of four basic hardware elements described below and illustrated in Figure 1 (for more information see, e.g., Corbin, 1981:7-14; Boss, 1979:9-13)

Central Processor Center: This component, often called the CPU, consists of a main memory unit of high-speed storage to hold the immediate information being processed and the step-by-step instructions for processing the information, a control unit to interpret the instructions and coordinate operations to carry out the instructions and control the other components, and the arithmetic-logic unit for adding, subtracting, dividing and comparing information following the set of instructions. Each of these components is made up of one or more chips, which are integrated circuits of thousands of tiny electronic parts like transistors and resistors on areas often no larger than a square centimeter (see Griffiths, 1981:5-15 for information on changing information processing capabilities and costs).

Auxiliary storage: Since the amount of main memory is limited, computers usually have auxiliary storage devices that use magnetic disks or tapes that have longer access times than main memory, but which can store much larger amounts of data at much lower cost.

Input devices: Information (instructions or data) are submitted to the computer for processing from typewriter terminals, keyboards punched cards, optical character readers, magnetic tape and disc drives, etc.

Output devices: Information is obtained from the computer through video displays, typewriter terminals, high speed printers, etc. The output from the computer may also be transferred to magnetic discs or tapes.

The latter three components are peripheral devices or peripheral equipment and the CPU with the amount of main memory plus the peripheral equipment is called the computer configuration of the particular installation.

Figure 1 around here.
(Simplified Diagram of a Computer.)

Computer memory, whether high-speed main memory or auxiliary storage such as a magnetic disk is stated in term of kilobytes, e.g., 64K bytes, where the K stands for approximately 1000 bytes, or to be exact 1,024 bytes. Each byte normally represents an alphanumeric character (i.e., a letter, number or special character). A byte is a string of bits, which is the basic unit utilized by a digital computer and have values of 0 or 1. A combination of bytes (and therefore of bits) makes up a computer computer word, usually 8, 16 or 32 bits long.

Bibliographic information is usually stored on magnetic disks because this form of storage allows almost immediate access to any part of the information in the same way that the analogous phonograph record allows one to go directly to a particular song by moving the phonograph arm. Information stored on magnetic tape, on the other hand, is a much cheaper form of mass storage but it does not allow direct access, since like on an ordinary tape recorder, the tape must be advanced or rewound to arrive at any particular place on the tape.

2.A.2 Software

Computer hardware can do nothing without receiving instructions. The collection of computer instructions to drive the computer is known as a computer program or quite frequently as computer software. There are three important classes of software:

Computer system software, usually provided by the hardware manufacturer, includes the operating system, which controls the overall operation of the computer, and compilers and interpreters that translate computer programs written in high-level programming languages that are "English-like" such as BASIC, COBOL, FORTRAN, etc., into the machine language understood by the computer.

Generalized program packages are systems that are specific to a particular problem area but general in that many different problems of the same class can be treated. For example, the widely used Statistical Package for the Social Sciences (SPSS) can be made to produce many different tables and statistics from the micro-data of any population survey; similarly, a generalized bibliographic storage and retrieval package like ISIS (see Section 7.D.1) can be used for producing an abstract journal or carrying out a search.

Application programs, which are the instructions that direct a generalized program package to carry out a particular task on a specific set of data. For example, the ISIS package has to have an application program telling it what information to include in an abstract journal, how it should appear on the output page, etc. A program written by a programmer for a special purpose (such as to calculate a checkbook) can also be called an application program.

Many generalized program packages can be obtained as written by the programmer (the source deck) or as translated (compiled program) by the computer into machine language (the object deck). To allow a program to be adapted to some special need of the user or to correct an error (bug) found in a program, the source deck must be available, although suppliers of software often provide only object decks. If the supplier does not agree to correct bugs found in the software (known as software maintenance for which there may be an annual fee), the source deck is essential and the user must have programming assistance available when needed.

2.B COMMUNICATION WITH THE COMPUTER

Although the input and output of information to and from the computer may be done through IBM cards, it is increasingly common to communicate with the computer via video or typewriter terminals. As the terminal sends and receives messages from the computer over a communications line, this is known as working on-line. In addition, if the user and the computer can exchange questions and answers during, for instance, a search session, then the user is working in an interactive mode. In batch mode, on the other hand, the user must give the computer all the information for the complete search in a single batch and will receive the results when they are ready; any change to be made in the formation of the search requires starting all over.

If the system is on-line, the input and output devices can be located at any distance from the computer, usually linked through telephone lines. Telecommunication requires that the signals from the computer be transformed by a modem into signals able to be transmitted by telephone lines or other link and then retransformed back into computer signals by another modem at the point of reception where the user's terminal is located (see Figure 2).

Figure 2 around here.
(Telecommunication links with a computer.)

More than one input and/or output device can be used with a computer at the "same" time, via some form of time-sharing that permits the computer processing time to be shared among several completely independent users; since the computer spends only a fraction of a second with a user before attending to the next, each user is unaware of the others, although if there are many users the computer may have a slow response time.

2.C CATEGORIES OF COMPUTER

For many purposes, including the present Guide, it is convenient to divide computers into three general size categories: mainframe, mini- and micro-computers. The definitions, based on CPU processing speed, memory capacity, cost, physical size, etc, are relative, since the capacities of mainframe computers 15 years ago are now those of some micro-computers. For purposes here, the definition of a mini-computer given by Griffiths (1981:23) will be followed: "a small, general purpose computer that does not require carefully controlled environmental conditions for operation, and which can be used in a variety of applications". Depending on the peripheral devices purchased with a mini-computer CPU, the cost can range from say US\$20,000 to US\$150,000. Mainframe computers, like the IBM 370 and 4300 series cost above US\$150,000, have much more stringent site conditions, and unlike mini-computers, normally require trained operators and programmers specialized in their operating systems.

A micro-computer is a very small computer, normally consisting of a keyboard and some form of video monitor, all of which is not much larger than a traditional office typewriter; the CPU is a micro-processor on a single chip. Auxiliary storage is usually a diskette or an ordinary tape cassette and many micro-computers designed for personal use in the home utilize an ordinary television set for their output display (for more information on micro-computers, see, for example, Griffiths, 1981; and Toong and Gupta, 1982). Depending on their CPU capacity, memory, peripherals, etc., the cost of a micro-computer in 1982 ranged from US\$80 to around US\$40,000. In the near future, it seems likely that powerful micro-computers with suitable auxiliary memory and software will meet the bibliographic retrieval needs of many PI Centers in the developing countries (see Section 7.D.2).

3. COMPUTER STORAGE AND RETRIEVAL OF BIBLIOGRAPHIC INFORMATION

To understand the basic principles of the computer storage and retrieval of bibliographic information it is not necessary to understand the actual way in which the computer operates or how the information is physically stored and

moved about by the computer. However, it is desirable to understand the logic of what the computer can do for the user.

3.A CREATION OF SECONDARY BIBLIOGRAPHIC INFORMATION

In general, to retrieve one or more documents from a set of many documents, whether using computerized or manual procedures, requires that new information succinctly describing each document be created and placed on "cards" for easy manipulation in, say, a card catalog or computer. This secondary information, or more specifically bibliographic information, on each document is organized in a convenient manner, such as by subject, to permit the location of the relevant "cards" in response to user questions; with the information given on a retrieved "card", the user can locate the physical document. In a computerized system the "card" or "cards" describing each document are called bibliographic records or, in general, records, and a set of records treated together is called a file; when IBM cards are used, the file is often called a deck.

3.B COMPUTER RETRIEVAL OF BIBLIOGRAPHIC INFORMATION

How is the secondary information describing a set of documents organized so that the computer can locate the master record (or main record) that fully describes each document from the complete collection of master records (known the master file or the main file) that will be searched together? Each master record has a unique record identification number to distinguish it from other records in the file and a description of the identifying characteristics (title, author, etc.) and the content of the document including keywords from a controlled vocabulary. When the controlled vocabulary is a thesaurus, the keywords are called descriptors.

Each item of information about the document is normally stored in separately identifiable sections of the record known as data fields or simply as fields. Usually each field has a numeric or alphanumeric tag or field number such as A05 or 25. See Figure 3.

Figure 3 around here.
(Simplified diagram of a computer record and its fields.)

Two general approaches may be used to carry out a computerized search. In a serial search (sometimes called a text search), the master records of the database are read one by one by the computer to locate those that have the terms specified in the user's search, which might include author names, words from the title, keywords, the abstract or any other field. Obviously the computer must read the entire file (or all the relevant fields of each record) to insure that all records of interest are retrieved. See Figure 4. This procedure will be slow and inefficient, just as it would be for a person to read an entire book on urbanization to determine on what pages the author speaks of the rank-size rule.

Figure 4 around here.
(Serial and inverted file searches.)

However, just as one can create an index for a book to allow the user to go directly to the letter "r" to locate the rank-size rule, a similar approach known as an inverted file search can be employed by the computer. An inverted file has a special record for each term; the record contains the identification numbers of all the master records with the term. Consequently, when the computer is asked to find, for example, all documents with information with information on both Peru and Colombia, rather than read all the documents, it can go directly to the inverted file records on Peru and on Colombia to compare the identification numbers of the documents, selecting only those that are present on both of these inverted file records (see Figure 4). Annex 2 explains the logical operations - Boolean Algebra - that are carried out to define and combine sets of records during a search. With the identification numbers, it can locate and print the master record of each of the documents.

Often in computerized systems the master file along with the inverted file as well as any other associated files, are called a bibliographic database or simply a database. Note that in computerized systems at present, the actual documents are NOT in the bibliographic database; rather the database is the secondary information describing the documents. RECORDS, NOT DOCUMENTS, ARE RETRIEVED BY THE COMPUTER.

4. ADVANTAGES AND DISADVANTAGES OF COMPUTERIZATION

The major advantages of computerizing population information storage and retrieval in a PI Center are:

a). Users have vastly increased capabilities to find specific information because more complex combinations can be searched, more access points are possible (the information in any field of a bibliographic database can be setup for searching) and any number of terms can be included in the search -- for example:

Locate all documents, including book chapters, that use World Fertility Survey data for comparative studies of the effect of infant mortality on fertility in two or more of the countries Peru, Ecuador, Colombia, Panama, Paraguay or Costa Rica and that discuss the policy implications of the findings. The documents should be in Spanish or English, published after 1980, and exclude any documents presented at the 1981 IUSSP General Conference or journal articles written by the author, J. Perez (the logical operations that must be combined to carry out such a search are outlined in Annex 2).

b). The quality of the secondary information can be increased and the products derived from it can be improved since if it is entered correctly it can be alphabetically ordered and reproduced by computer whenever needed;

c). Many routine and error prone manual tasks are eliminated like alphabetizing catalog cards or re-typing citations for users or lists;

d). The bibliographic database can be utilized for new services at little extra cost and often without adding new permanent staff (e.g., an

abstract journal can be produced from a database planned primarily for searching);

e). It is easier to deal with fugitive literature which is difficult to describe and retrieve manually; and,

f). Exchange of information with other centers is facilitated.

However, in a small PI Center specializing in population literature, there may be significant disadvantages that may outweigh the advantages in some circumstances. Some of these disadvantages are:

a). The work of the PI Center may become dependent on the computer in situations when the operation of the computer is not reliable or there are long delays between a request for a search and the receipt of the actual bibliography. Furthermore, whenever the computer is down (down-time), because it has crashed due to a failure or is halted for preventative maintenance, neither the users nor the documentalists may have a way to obtain the information they require;

b). All communication with the computer must be done by programs. If errors are found or new functions are required, the documentalists normally have to obtain the services of a programmer to solve the problem;

c). Greater precision in indexing and other input may be required than in a manual system, simply because in a manual system the errors would not be noticed; similarly, the computerized system will probably reveal that many documents entering the system are of little interest to users - the existence of many of these would not have been noted in a manual system;

d). The staff must be re-trained to perform new tasks and learn new ways of thinking about information;

e). Users of the PI Center may have to be "re-trained" to use new services, such as on-line searches that replace older card catalogs; and,

d). Initial costs are likely to be relatively high and there probably will not be a decrease in the overall cost although more work may become possible with the same staff (Boss, 1979:3). Some costs will rise, e.g., paper, as many lists not previously possible will be routinely printed.

5. COMPUTERIZATION OPTIONS

Assuming that there is no reliable local service available to perform the computerized operations for the PI Center at a reasonable cost, four general alternatives are open to the manager:

1). Computerize as soon as possible;

2). Continue with the present manual system because it provides good service to users and is likely to do so for the foreseeable future (converting to a computerized system would have high initial costs not justifiable for the volume of documents or users or the types of services

required);

3). Utilize a manual system which employs modern retrieval techniques to improve user services and which allows easy and rapid future computerization. This should be the option selected by new PI Centers that do not have computer access, funds or present need for computerization; and,

4). Use the computerized search facilities of databases elsewhere.

The fourth option, which allows any PI Center to obtain the benefits of computerized searches, is clearly not in conflict with the other alternatives and should be considered whenever appropriate even in computerized PI Centers.

6. CRITERIA FOR EVALUATING THE CONDITIONS AND DESIGN REQUIREMENTS FOR COMPUTERIZATION

The decision concerning which option to follow and how to design a computerized system should be consistent with the characteristics and needs of the PI Center's users. Answering the questions discussed below will help the manager to determine the appropriate option and to define key design requirements if computerization is selected.

6.A HOW MANY DOCUMENTS?

While it is difficult to set an absolute figure on the minimum number of documents to warrant computerization - in part it will depend on the types of users, the specificity of the information required and the output products -, it is evident that it would be difficult to justify computerization of a very small collection with a reduced number of new documents entering per year. It is likely that the service will be more reliable, more rapid and far less costly to operate manually than with presently available computer systems. If the existing system does not work well, better manual retrieval methods should be considered (Option 3 then would seem most appropriate).

The number of documents to be entered initially into the population information database and the rate of increase per year will depend both on the scope of the database and the definition of the item that is described on the bibliographic record (see Section 7 for explanations of scope definition, bibliographic level and documentary unit). However, as a rough figure, one might say that a library collection of a minimum of 1500 items with an increase of at least 400 or 500 per year is required to justify computerization if other conditions listed below are met.

6.B WHO ARE THE USERS?

If the users are primarily secondary school students or the general public concerned with very general questions - "what is available on population policy in Asia?" -, the potentialities of a computerized system will be very much under-utilized and unjustified.

If the users are primarily policy makers, high level administrators and similar persons, they normally require consolidated information; it is unlikely that they will utilize the database retrieval facilities and it is probably

unnecessary for the consolidators in the PI Center as well, unless the amount of information to be consolidated is large and varied.

On the other hand, if the users are analysts, investigators, etc., as might be the situation in a university, research center or national population information center, the need for specific information will be much greater and likely to increase with the experience of using a computerized system.

6.C HOW MANY USERS ARE THERE AND WHERE ARE THEY?

Even if the users require specialized information, the complexity of creating a computerized system will not be justified (with present methods and computers) if the annual number of requests is quite small, say no more than 50 to 100 per year and unlikely to increase because only internal users are served.

If the number of user requests is high enough to justify computerization, the physical location of the users - in the institution or the same city vs. elsewhere - will have an effect on the design of the computerized system. If they are local users, the benefits of interactive searches can be maximized since the search results can be obtained immediately and the user can do the search him or herself or work with the documentalist, redefining the topic as the problem becomes clearer from seeing what is available. If the users are external, requesting services by mail, batch searches may be acceptable, although the PI Center searchers will find their work much easier to do better if the system is interactive.

6.D WHAT SERVICES ARE REQUIRED?

If the computerized system is justified by the volume or the specialized needs or characteristics of the users, the types of services that will be provided will affect the design of the system. The major services that should be considered for provision from the computerized database are:

Searches	Selective dissemination
Abstract journal	Standardized updated bibliographies
Internal lists and indices	on subjects of wide interest

Most PI Centers should design their computerized system to be able to produce searches, lists and indices and an abstract journal. It is important to note that a system designed for interactive searches using computer programs oriented primarily toward searches (such as the IBM program STAIRS) may not find it easy to make lists and to produce an abstract journal as an additional service.

6.E WHAT "COMPUTER ENVIRONMENT" IS REQUIRED?

A manual system may have many limitations, but it can be reliable even if the physical conditions are unpleasant and variable. On the other hand, most of the computers presently used for bibliographic databases require conditions that may not be easily satisfied over long periods of time such as a source of electrical energy with minimal fluctuations in voltage, well-filtered air-conditioning to control dust and temperature, periodic preventive maintenance (when the computer will not be available) and rapid service if the computer goes "down" (down-time) and has to be repaired. If telephone lines are to be used to communicate with the computer, they must be tested before taking any decisions to insure that the

signal transmission quality is adequate over long periods.

If the PI Center or its parent agency is going to obtain a computer, rather than use computer services provided by another agency, it is extremely important that there be adequate hardware maintenance and operating system maintenance by the computer equipment manufacturer's authorized service; if service is not readily available or is unreliable, and there is not a back-up computer, i.e., an immediately available substitute computer with the necessary characteristics, computerization should not be considered until a suitable computer is available. Even if only terminals, printers or modems are in the PI Center, they also will require periodic preventative maintenance and repair of failures by an authorized service with access to spare parts.

6.F WHAT RESOURCES ARE REQUIRED?

Human resources: In most cases, computerization will require retraining of staff accustomed to think in terms of more traditional library or documentation procedures and it may require psychological re-orientation since the pace of work and the accuracy required will be much increased.

It is likely that experienced outside consultants, such as systems analysts, will be required to assist in making some of the more technical decisions. Unless a turn-key system generalized program package is employed, i.e., it is ready to be used by figuratively just "turning the key", programming assistance will probably be required to install the software and to write the application programs. In such a case, if a programmer is not available for a large percentage of the time in the first year and is not readily available when problems come up thereafter, the service may be very unreliable and/or inflexible. Furthermore, if there is only one programmer who knows the technical details of the system, there may be problems when he or she is on vacation or leaves for another position; consequently, it is highly desirable that a second programmer have some technical knowledge of the system and the application programs.

Financial resources: It is not possible to provide useful figures on the cost of computerizing since they will depend on local costs for staff, services and equipment as well as local factors such as whether the parent agency has a computer that can be used. However, it is likely that the initial costs of creating the computerized system will be a significant percentage of the total annual cost of operating the manual system of a small PI Center and may even exceed it in some cases; that is, the overall cost of the first year or two will be higher than previous annual costs. Furthermore, it is unlikely that the normal annual cost of maintaining the computerized PI Center will be less than when it was manual; rather, as indicated in the advantages of computerization, the service should be improved and more services can be made available.

Although the cost items for starting up and continued operation will depend on how the system is designed, Tables 1A and 1B should help the manager to identify all possible costs to avoid ignoring an item during the planning stages (see also Corbin, 1981:50-58).

Table 1A around here.
(System Development and Database Creation Costs.)

Table 1B around here.
(Operational Costs.)

7. BASIC DECISIONS IN THE DESIGN OF A COMPUTERIZED SYSTEM

It will be assumed in this section that the decision has been made to create a computerized database and that the services derived from the database will be searches on request and the production of a journal (see UNESCO, 1980b for a guide on using a computerized database for the selective dissemination of information). A number of key system design decisions must now be made on substantive, documentation and computer matters. It is important to remember that these are highly inter-related and that many of the substantive and documentation decisions will depend ultimately on the software that can be obtained (at a reasonable price) for the computer equipment to which the PI Center has access.

Although computerization will no doubt become easier in the future if inexpensive turn-key systems using micro-computers are developed, at present, the establishment of a computerized database and the operation of services based on it are relatively complex and costly. For this reason, the manager of a small PI Center in a parent agency with limited resources should take the following two points of advice very seriously:

1). USE INTERNATIONALLY RECOMMENDED PROCEDURES ADAPTED TO MEET THE NEEDS OF THE PI CENTER - DO NOT TRY TO INVENT NEW SYSTEMS. Concentrate instead on ADAPTING AN EXISTING SYSTEM FROM ANOTHER CENTER to the specific requirements of the PI Center.

2). OBTAIN EXPERT ADVICE from systems analysts and consultants on computer equipment, bibliographic information processing software and the application of these to the specific PI Center's needs and environment.

7.A FLOW DIAGRAMS OF THE INPUT AND OUTPUT OPERATIONS

The flow diagrams in Annex 1 show the basic operations that a PI Center will be likely to perform to enter (input processing) information on documents into a bibliographic database and to obtain output (searches and journal production). The operations are depicted in very general terms to apply to a wide variety of situations. To clarify the actual design of the PI Center's own system, the manager should make similar diagrams for his or her own PI Center, breaking down each of the boxes into smaller individual steps.

7.B DECISIONS CONCERNING INPUT

7.B.1 Scope definition

All bibliographic systems should have a scope definition that indicates which documents should be included and which should be excluded. A carefully written definition should state the subject matter, type of bibliographic entities that will be included, earliest date of publication, geographical coverage and the languages accepted. This is particularly important for a computerized database since items that do not "belong" will be much more evident to users in

a computerized system than in a manual one. Furthermore, it is costly to go through the entire input procedure if it will never be consciously requested because it is not expected to be there.

It will be assumed here that the type of bibliographic entity included will be any published or unpublished (fugitive literature) item that is a serial, monograph, conference document, report or thesis or any part of these.

As there are regional population documentation systems in many of the developing regions, most PI Centers will want to limit their geographical coverage to their own country, relying on the regional centers to provide national users with searches from the rest of the same region. Ideally, at least one national center should provide its country's input to the regional center, a relationship that will be facilitated if they both use compatible input formats; compatibility implies that a computer program could be written to transform the records from one format to the other.

7.B.2 Bibliographic levels and documentary units

Since the computer facilitates the location of specific information, it is desirable to divide a larger document into smaller units, with a master record for each such unit. The decision to do this, which is highly recommended, will have an important effect on the number of bibliographic records in the database since documents such as books will generate a record for each relevant chapter.

To clarify what is involved and point toward how the smaller parts of documents can be treated in a computerized database, it is convenient to adopt the concept of bibliographic level. Following the "Reference Manual for Machine-Readable Descriptions" (UNESCO RM, 1981), a computer record can describe any one of the following four bibliographic levels (see UNESCO RM, 1981:1.3.1-1.3.3 or UNESCO, 1980a:172-173):

(a). The analytic level which describes an item that is part of a larger work and whose description cannot stand alone (e.g., a chapter of a book, or an article from a journal);

(b). The monographic level which describes an item whose description can stand alone (e.g., a book, thesis or conference pre-print);

(c). The collective level which describes an item that is composed of two or more physical pieces in a set with a definite number of constituent pieces (e.g., the set of manuals from a survey);

(d). The serial level which describes an item issued in successive parts intended to be continued indefinitely (e.g., a journal, monographic series or annual report).

The item being described in a bibliographic record, such as a chapter from a book, is defined as the documentary unit and the whole from which it is taken is the generic document (Morin-Labatut y Sly:1982:11, 27-31). Of course, if a book as a whole is also described on a separate record, the generic document is the documentary unit for that record. Consequently, when estimating the amount of auxiliary storage that will be necessary for the database (see the Section on

Hardware), the number of records in the database will be the number of documentary units, not the number of physical documents in the collection.

7.B.3 Bibliographic record content

The content of the bibliographic record and its format involve a series of decisions which will have a major impact on the quality and type of services that can be provided from the database, the possibilities to exchange information in the database with other centers and the cost of creating and entering the input. Four types of information are normally included: (a) a bibliographic description, including the title, author, etc., intended to provide a unique and unambiguous reference to the document; (b) an abstract providing a representation of the content; (c) indexing terms, such as descriptors, to describe subject content with terms from a controlled vocabulary; and, (d) local information of interest only to the PI Center, e.g., record entry date, documentalist name, etc.

7.B.4 Bibliographic description

The bibliographic description is made up of data elements which are pieces of information describing the document, such as the title, author name, page numbers, etc., which are separately stored in the data fields of the bibliographic record and individually identified by alphanumerical field tags. Storing each element separately is extremely important in a computerized system because this allows the elements to be independently identified and easily manipulated by computer programs; such manipulation is not possible if various data elements are grouped together as on a library catalog card (UNESCO RM, 1981:1.1.2).

7.B.5 Selection of international and internal formats for bibliographic description

The PI Center must decide on the format of its bibliographic description and other fields. The format comprises (a) the record structure (known as the carrier format) including the layout of the data fields; (b) the data element definitions; and (c) the specification of the data element sets which are required for each combination of type of bibliographic entity and bibliographic level, e.g., the data elements necessary to describe a chapter from a monographic series. The data element definitions explain how the information in each field is to be entered, such as how the date of publication must be written to permit easy recovery by computer.

To facilitate exchange of bibliographic information among centers there are international exchange or communication formats. The two major internationally agreed upon formats for general use are: the UNISIST Reference Manual format (or Reference Manual format) originally oriented toward abstract and index services, and the UNIMARC format, designed for interchange among libraries (for a summary of the basic characteristics of these and other formats see UNESCO, 1980a:169-179).

As the PI Center in a developing region will wish to exchange (or at least receive) bibliographical data from the regional population information center working in coordination with POPIN, the PI Center's internal format for its own

bibliographic record should have compatibility with that used by the regional center, which in turn must use a format compatible with an international format. Although at the time of writing only one of the developing regions has a computerized regional population documentation system (the CELADE/DOCPAL Latin American Population Documentation System located in the United Nations Latin American Demographic Center, CELADE), it is very likely that all the regional population information centers being located in United Nations agencies, will be compatible with the UNISIST Reference Manual format.

Consequently, if the regional population information center does not yet have a computerized system, it is recommended that:

THE PI CENTER USE AN INTERNAL BIBLIOGRAPHIC RECORD FORMAT BASED ON THE UNISIST REFERENCE MANUAL FORMAT (see next section).

For an inexperienced PI Center, the Reference Manual format also has the advantage over, for example, the more complex UNIMARC format in that the Reference Manual is a self-contained manual providing the cataloguer with all the information needed in a single document (Dierickx, 1982:24-25). Naturally, if the PI Center is part of a larger information system, such as a university library, it also may have to adjust to the norms of its parent body.

7.B.6 The internal bibliographic record format and the PI Center Worksheets

As seen in the input flow diagram (Annex 1B), before entering the bibliographic information of a documentary unit into the computer, its bibliographic description, abstract, descriptors, etc., must be placed on worksheets, which contain spaces with their respective field numbers indicated to write the information for each data element that is included in the PI Center's own internal format. Of course, the specific set of fields used (data element set) on a worksheet for an actual documentary unit will depend on its bibliographic level and the type of bibliographic entity.

Annex 3A shows the worksheet now utilized by CELADE/DOCPAL and designed in collaboration with the Economic Commission for Latin America (CEPAL) in order to have a common worksheet for all economic and social development bibliographic databases in CEPAL and to serve as a model for national centers in the Latin American region. Associated with this worksheet, as with any other is a Field Definition Table (not shown here) that shows for each field the field number, extended title, maximum number of characters in the field and the characteristics, such as whether the field has a fixed length or can be any number of characters up to the maximum and whether the field can be repeated (e.g., to enter each author's name, if more than one). The manual (CEPAL, 1982) that describes how each field is entered, is compatible (compatibility) with the UNISIST Reference Manual format so that a computer program can be written to transform the records with the internal CELADE/DOCPAL format into the Reference Manual format.

Annex 3B shows the worksheet designed by the International Development Research Centre (IDRC) of Canada for development-Information systems. It is somewhat further removed from the approach taken in the Reference Manual than the worksheet used by CELADE/DOCPAL, since the former is also designed to be compatible with the approach taken by the International Information System for

the Agricultural Sciences and Technology, known as AGRIS; however, the records can also be transformed by computer program into the Reference Manual format. The IDRC manual (Morin-Labatut and Sly, 1982) and the CEPAL manual described above both adhere to the ISO standards (ISO, 1977) for codes for countries, languages, etc.

To facilitate the exchange of information with its regional population information center and to avoid errors and duplication of effort the following recommendation should be followed:

The PI CENTER SHOULD USE THE WORKSHEET AND ASSOCIATED PROCEDURES MANUAL EMPLOYED BY THE REGIONAL POPULATION INFORMATION CENTER, adding additional local fields if necessary and removing any unused local fields of the regional center.

If the regional center does not yet have a worksheet for a computerized system, a worksheet from another region following the norms outlined here should be used in coordination with the regional center.

The manager should include on the worksheet only the data elements that are absolutely necessary for the PI Center or which are defined as essential or mandatory fields by the regional center manual. This policy will reduce costs, diminish errors due to over-complexity of the worksheet, simplify the training of input processing staff and help speed-up the input processing.

7.B.7 Abstracting and indexing

If a journal is to be produced or searches will be sent to users outside the PI Center, it is highly recommended that an abstract with substantive information be included since it is unlikely that the users will have easy access to the actual documents, many of which will be unpublished. Computer printouts of the abstract may satisfy their needs directly or at least indicate if it is worth trying to obtain the document. However, if this decision is taken, part-time workers or staff able to write suitable abstracts must be found and their costs included in the operating budget. Furthermore, the software and auxiliary storage (normally magnetic disc) must be able to handle a record size that includes an abstract with up to as many as 2000 characters (300 words, although normally it will be considerably shorter).

The documentary units should be indexed with keywords even if an abstract is included, because the software available to facilitate searches of abstracts is costly and not likely to be available to small PI Centers or their parent agencies (text searching in commercial databases like DIALOGUE is described in Meadow and Cochrane, 1981:94-106).

The "Population Multilingual Thesaurus" (PMT) (CICRED, 1979), now available in English, French, Spanish and Portuguese versions, should be the source of descriptors for the PI Center. This will allow automatic computer "translation" into the other PMT languages if that is necessary later. When a descriptor is not available in the PMT, the OECD Macrothesaurus (OECD, 1979) and the POPLINE Thesaurus (1981) should be utilized. See the Report of the 1982 POPIN Advisory Committee Meeting (POPIN, 1982) which includes rules for harmonizing the use of the PMT with the other two thesauri.

7.B.8 Data entry, correction and modification

After completing a set of worksheets, the information on each worksheet must be entered into the computer along with the corresponding record identification number. It may be possible in some cases for the PI Center to send its worksheets to an efficient central processing center and receive back various output listings and/or a magnetic tape with the latest updated master and inverted files so that the PI Center can make searches without being involved in other computer matters. It might even be possible to connect by telephone line to the processing center so that searches can be conducted by the PI Center. However, it will be very difficult to provide adequate services if the worksheets must be sent outside the local area for processing because of mail delays and difficulties of communication over long distances. For purposes here, it will be assumed that the PI Center has to enter and correct its own bibliographic data.

The data entry can be done in batch mode or interactive mode. In batch mode the data are often entered off-line into a file and when a large enough batch is accumulated, it is entered into the master file of the database. A well designed interactive system is likely to be somewhat slower but it will identify many errors that can be corrected immediately, saving correction time later. The decision on which approach to use will depend on the software available and the costs (staff vs. machine) involved in the local situation.

The software should permit both individual record correction and global data correction, in which all records with a the same error or requiring the same modification can be changed together. Particularly if a journal is to be produced, it is highly desirable to have software that checks for all possible logical errors, such as a date of publication that is later than the data entry date, as well as errors in standard terms such the use of an illegal or misspelled descriptors.

7.B.9 Inverted file

As seen in Section 3, it is highly desirable to be able to make searches on an inverted file. In some widely used storage and retrieval software, such as the IBM versions of ISIS, new or corrected bibliographic records entered into the master file are not automatically included in the inverted file. Rather the selected fields from each record of the whole master file must be re-inverted even if only one new record is to be added to the inverted file. If the database has many records, this can be very costly. Consequently, it is desirable to use a storage and retrieval software system that has some form of automatic inversion.

7.B.10 Database security

The accidental destruction of a bibliographic database due to physical damage or the introduction of an error because of a programming fault is a catastrophe that is certain to be too costly to repair if proper measures of database security are not included in the design of the overall system. While the actual procedures to be followed are beyond the scope of this Guide, they involve periodically copying of all the files associated with the database onto magnetic tape or other auxiliary storage medium and keeping these back-up copies from a fixed number of previous back-up cycles, normally in another physical

location so that a disaster like a fire does not destroy both the original and the back-up copies. It is necessary to keep tapes from various back-up cycles since that helps to recover a database if it is discovered that a recent change has accidentally introduced serious errors into the database.

7.C DECISIONS CONCERNING OUTPUT

7.C.1 Searches

From the discussion in Section 3, the description of Boolean Algebra for searches in Annex 2 and the flow diagram in Annex 1C, it should be evident that better service can be given with interactive searches than with batch searches, since the searcher can "learn" from an examination of the information being retrieved and make modifications to come closer to the user's needs. Consequently, it is desirable to obtain interactive search software, although batch searches for making large standard lists, etc., should also be possible. The software should be user-friendly, in that it should be easy for a documentalist or a motivated user to learn and should not require assistance from programmers. For additional comments that affect software decisions in this area of output, see the Section 7.D.

7.C.2 Production of journals and listings

The production of listings for internal PI Center purposes or an abstract journal will require that the storage and retrieval software be able to generate listings of citations, etc., with the items sorted on one or more selected fields such as the descriptor, publication date, and title, and may require that applications programs be written to re-organize the output into the final form desired.

The programming of complex listings will probably require the services of a programmer: applications programs for frequently required listings should be able to be routinely generated by simple commands from a documentalist. Whether this is possible will depend on how user-friendly the software and computer are.

If a journal is to be produced, the manager must take a series of decisions on the physical layout of the pages, the design of the indices, etc. Following the general philosophy expressed in this Guide, if the regional population information center has a journal format that meets most of the needs of the PI Center, the regional center's journal format should be adapted to the PI Center's requirements. This may also allow the PI Center to use the relevant applications software of the regional center. It should be noted that it may be necessary to obtain special text formatting software to control the final output, such as double columns, changes in letter sizes, page numbering and headings, etc. See Section 7.D.3 for aspects of printer selection.

7.D DECISIONS CONCERNING HARDWARE AND SOFTWARE

The interrelated decisions concerning computer hardware and bibliographic storage and retrieval software are highly technical and will have very important effects on the development and long-term operating costs. Furthermore, decisions made at the beginning may be difficult to change as more experience is gained. Finally, computer hardware costs are falling very rapidly while capa-

bilities of small machines are increasing. Consequently, this guide can only provide an overview of present likely computer/software combinations and suggest possible advances that may occur in the next few years which could radically alter the possibilities for the computerization of small PI Centers. For these reasons, it is imperative that EXPERT ADVICE BE SOUGHT ON COMPUTER/SOFTWARE COMBINATIONS APPROPRIATE TO THE NEEDS AND CONDITIONS OF THE PI CENTER AND ITS LOCAL ENVIRONMENT.

It will be assumed that most small PI Centers in developing countries will want to conduct interactive searches and produce listings and/or an abstract journal but do not have an agency in their local area to provide computerized bibliographic database input and output services (or that alternative is too costly or unreliable). The PI Center then has two approaches to obtain a storage and retrieval software package meeting the needs outlined in the previous sections of this Guide:

(a). Obtain a suitable software package from another agency (ideally a turn-key system), adapting it if necessary; or

(b). Develop its own software package "in-house".

A small PI Center should NOT consider the second alternative unless there is no software available for the computer to which it has access, and then only if it has had extensive experience with computerized bibliographic databases and with computer software design and programming and can tolerate long delays (see Section 7.D.2 for future solutions that may soon be available).

7.D.1 Computer/software combinations available

Although the PI Center will want to select the computer and software independently of each other in order to get the features required in each at the lowest costs, most small PI Centers in developing countries will have very little choice since the software systems available each must be used with the computer of a specific manufacturer (i.e., they are not portable software, able to be used on computers of different manufacturers). The PI Center may be constrained to the computer that is available in the parent agency or to which suitable telephone connection can be made for service at a reasonable cost; even if a computer can be purchased, leased or rented, the make, model, and operating system will be limited by the existence of adequate hardware maintenance (see Section 6.E on computer communication and environment requirements).

Furthermore, the PI Center will probably be limited to non-commercial software for the mini- and mainframe computers available, since suitable commercial software systems are extremely expensive to purchase or rent (see Section 2.A on other relevant software considerations).

There are two major non-commercial software systems available, ISIS, and MINISIS, for IBM mainframe and Hewlett-Packard HP-3000 series mini-computers, respectively. These, and a number of other computers with software available to permit on-line retrieval, are summarized in Table 2 (see also Griffiths, 1981:36-38 for commercial systems for mini-computers). It should be noted that not all the sources shown for the software may be able to provide software

maintenance, i.e., helping to install it and correcting errors that PI Centers encounter, and not all the software is very user-friendly, even to programmers. If maintenance is not guaranteed, it is essential to try to obtain the source deck (see Section 2.A.2).

Table 2 around here.
(Important Computer/Software (non-commercial) Combinations
for Bibliographic Information Storage and Retrieval.)

Valatin (1981) gives a detailed comparison of the capabilities of CDS/ISIS, the "official" version of ISIS which is fully maintained by UNESCO-Paris, and MINISIS, which is fully maintained by IDRC-Ottawa. These are probably the preferable software systems to use at present, but it is likely that many small PI Centers will find that buying time on an IBM mainframe to use CDS/ISIS is very expensive (unless the parent agency has its own multi-purpose IBM computer with the operating systems required) and that the purchase of the minimum-required configuration of a HP-3000 series mini-computer for MINISIS will cost over US\$50,000 and hardware maintenance may not be readily available in their area.

7.D.2 The micro-computer/software solution

The computer/software situation described above will leave many PI Centers with only the alternatives of postponing computerization until a suitable data-base input/output service is available locally or until portable software for micro-computers becomes available so that an inexpensive computer with locally provided maintenance can be utilized.

Fortunately, the steadily increasing capabilities of micro-computers and memory devices, coupled with the decreasing costs of the equipment and the wider markets for software products that low-cost equipment brings, are likely to make computerization financially feasible in the near future for many PI Centers with relatively small databases.

For example, a system known as IMD IV+V ^{1/} should be portable among many micro-, mini- and some mainframe computers; it is being written in the widely used high-level programming language, UCSD PASCAL ^{2/} by the Institut fuer Maschinelle Dokumentation (IMD) in Graz, Austria, a non-profit research and development corporation. Most of the software development work is done on PDP mini-computers (PDP 11/2 and PDP/23) and the first implementation of the software is on a Xerox 810 micro-computer (based on the Z80 micro-processor). Release 3 of the system, which will contain the full output processing functions, is expected to be ready by the end of 1983 (Park and Stearns, 1982). No doubt other agencies will also develop bibliographic storage and retrieval software for micro-computers.

^{1/} "IV+V" = Informations-Vermittlung und -Verarbeitung.

^{2/} UCSD PASCAL = University of California-San Diego PASCAL.

The cost of the software and the micro-computer equipment, probably including some form of magnetic disk as well as diskettes for back-up data storage and exchange, is difficult to estimate at present, but probably would be around US\$10,000 to 15,000. As the micro-computer, itself, is likely to cost only a few thousand dollars, it will often be possible to purchase more than one to facilitate the work and act as a back-up computer should one fail.

7.D.3 Peripheral devices

Normally the manager of a PI Center will have more freedom in the selection of input and output devices than in the computer to be used since the former are less expensive and often there are various manufacturers who make equipment that can be used with a given computer. While technical decisions concerning peripheral devices and the wisdom of using a mix of equipment from different manufacturers in the local conditions will require on-site expert advice, a number of factors should be kept in mind:

(a). A visual display unit (or video terminal) is the most rapid and convenient form of input and output in an on-line interactive retrieval system and may also serve for data entry (see Section 7.B.8); if retrieval services are to be available throughout the workday, it will be necessary to use separate terminals for input and output. Features of the display (number of lines per screen, number of characters per line, etc.) should be considered carefully as should the ease of using the terminal, since like software, a terminal should be user-friendly. Typewriter terminals for interactive work should be avoided.

(b). Since it will often be necessary to printout search results and applications programs on paper (hardcopy output), if only one printer can be obtained, careful consideration must be given to the speed and quality of the printed output, particularly if listings will also serve as camera-ready output for publication.

(c). The character set of both the input and output devices should have both upper and lower case characters since output written entirely in upper case is very tiring for documentalists and users to read and could tend to discourage acceptance of the computerized system. Any special characters, such as "=", "<" and ">", as well as any diacritical signs, accents, etc., required by the carrier language used on the bibliographic record, must also be available.

(d). The type of auxiliary storage, and the amount of storage in kilobytes will depend on the type of work being done (interactive and/or batch) and the design of the software and the specific files of the PI Center. In bibliographic files with abstracts, the average number of bytes per record (i.e., the record size of one documentary unit) may be around 3000 bytes (based on the CELADE/DOCPAL system).

(e). The cost of hardware maintenance (including preventive maintenance and repair of equipment failures), which normally costs around one percent of the purchase price per month (i.e., around 12 percent a year), must be included in the operating budget of the PI Center for all purchased equipment.

8. UTILIZATION OF A MANUAL SYSTEM TO FACILITATE FUTURE COMPUTERIZATION

From the discussions in the previous sections the manager of a PI Center may have come to the conclusion that it is either premature to computerize or that it is not possible at present because of the costs or unfavorable "computer environment". If the manager wishes, nonetheless, to improve the PI Center's information retrieval services, or if a new PI Center is being established, serious consideration should be given to the third "computerization option" given in Section 5: UTILIZE AN EFFICIENT MANUAL SYSTEM THAT EMPLOYS MODERN RETRIEVAL TECHNIQUES AND THAT FACILITATES FUTURE COMPUTERIZATION.

Because of space limitations, only the general principles of a manual computer-oriented system for the storage and retrieval of bibliographic information will be outlined here. The manual system, of course, should be based on the procedures and internal format of a fully computerized system. In the manual system, a form of worksheet, with the bibliographic information for a documentary unit placed in the appropriate "fields", is the equivalent of the master record. However, since the usual worksheet is too large to be manipulated easily by hand, it must be replaced by a smaller work card, which has a subset of only the most important bibliographic description and subject content fields from the normal worksheet. See Annex 3C for an example designed by CELADE/DOCPAL 1/; it is compatible with the UNISIST Reference Manual.

The set of workcards, ordered numerically according to the card record identification numbers is the equivalent of the computerized master file. Using a manual search procedure, once a master card is located, the user can examine the abstract and citation. As the master cards are ordered by number and therefore easily replaced, those of interest can be extracted temporarily from the file to have the citations and abstracts photocopied to avoid the errors and tedium of re-copying manually.

There are various ways of setting up the equivalent of an inverted file for searching; a variation of the UNITERM system is the simplest for a small PI Center that does not wish to invest in an optical coincidence device, such as a Termatex machine, that may become obsolete within a couple of years if inexpensive user-friendly micro-computer systems become available.

In the UNITERM procedure, each descriptor is placed on a separate UNITERM card which has 10 columns, each headed by a successive digit, 1,2,3,...; see the example in Annex 3C. On a card of a given descriptor, the four digit master card identification number of each document is written in the same numbered column as the last digit of the identification number, e.g., identification number "0085" is written in the column headed by a "5". To find the master cards of documents treating both Peru and Colombia one compares - visually - the UNITERM cards for each of these terms, looking for identical identification

1/ There are differences in the field numbers of the CELADE/DOCPAL worksheet and workcard shown in Annexes 3A and 3C. The latter was based on the worksheet used to enter over 15,000 records between 1976 and 1981. Since 1982, CELADE/DOCPAL employs the worksheet given in Annex 3A; the new workcard for manual systems should be ready shortly.

numbers on each card, aided by the fact that identical identification numbers have to be in the same column on each card. The logical operations possible are the same as shown in Annex 2. The PI Center can create a manual "inverted file" for each field of interest, such as authors, descriptors, geographical area treated, conferences, etc.

The same general rules suggested for computerized systems should be applied to the design of the internal format of the manual system. If the Regional Population Information Center does not have a computer-oriented manual or a computerized system, THE WORKCARD OF THE PI CENTER SHOULD BE DESIGNED TO BE COMPATIBLE WITH THE UNISIST REFERENCE MANUAL FORMAT, which has been designed for non-computerized as well as computerized systems (UNESCO RM, 1981: 0.4). In this case, a workcard from another region can be taken as the model.

If the Regional Population Center has a computer-oriented manual system or a computerized system, THE WORKCARD SHOULD BE THAT RECOMMENDED BY THE REGIONAL POPULATION INFORMATION CENTER.

Annex 3D shows the worksheet of the Population Information and Documentation System of Africa (PIDSA) located in the Regional Institute of Population (RIPS) in Ghana. It is used manually but is designed to facilitate computerization when a computer becomes available. PIDSA uses its system to generate the journal "PIDSA Abstracts".

9. UTILIZATION OF EXTERNAL COMPUTERIZED POPULATION DATABASE SEARCH SERVICES

Irrespective of whether the PI Center is computerized or not, the fourth option presented in Section 5 is available to all PI Centers: UTILIZE EXTERNAL COMPUTERIZED POPULATION DATABASE SEARCH SERVICES whenever complete bibliographies with abstracts are required or the PI Center's own database or search capabilities do not fully satisfy a particular user's needs. This should be the standard procedure when a national center is asked to provide information on other countries in its region since for reasons of efficiency its own coverage of a population topic will be primarily limited to its own country.

External search services can also help the PI Center to identify documents that it should acquire (e.g., those written in the developed countries about the PI Center's country) by periodically supplying the PI Center with updated listings of all new documents entering the external database that are within the scope definition of the PI Center.

At present there are three computerized population databases with international geographic coverage: POPLINE available on-line through the MEDLARS system of the U.S. National Library of Medicine; POPULATION BIBLIOGRAPHY of the Carolina Population Center of the University of North Carolina, available through the commercial database service DIALOGUE; and CELADE/DOCPAL of the UN Latin American Demographic Centre. The first two use English as the carrier language for the abstract, etc., and the last, Spanish.

POPLINE, with over 100,000 records in 1982 and updated monthly, is the largest system with world-wide coverage of published and some unpublished materials; it has particularly wide coverage of population-related bio-medical family plan-

ning literature. The database was created from databases maintained by the Population Information Program of John Hopkins University, the Center for Population and Family Health of Colombia University, and Population Index of Princeton University. Persons in developing countries can request searches free of charge on the POPLINE Literature Search Request Form shown in Annex 4A. An example of a search citation with its associated English language abstract is also given. Copies of the documents can also often be supplied if not available in the user's country.

The Carolina Population Center's POPULATION BIBLIOGRAPHY is one of the more than 400 databases (in 1981) that are available through the DIALOGUE on-line search services. The population database, with around 47,000 records in 1981 and updated bimonthly, costs US\$55 per hour of connection time for terminals in the United States and US\$0.10 per record printed on a listing (from the database description in Meadow and Cochrane, 1981:177).

Annex 4B gives the form to request searches free of charge from CELADE/DOCPAL on the population literature written in or about Latin America or the Caribbean since 1970; at present the database has around 17,500 records increasing at around 2000 per year. The abstracts are in Spanish, although titles are given in English and Spanish as well as in the original language. As many of the documents are unpublished and therefore difficult to obtain, all documents can be supplied on request, in some cases paying the cost of copying the pages (see the Document Request Form in Annex 4B).

For users working in population who wish to read documents from related fields, PI Centers should also consult relevant non-population data bases in the UN regional commissions and in the developed countries.

BIBLIOGRAPHY

- Boss, Richard W. 1979. The Library Manager's Guide to Automation. White Plains: Knowledge Industry Publications.
- CELADE. 1980. Procedimiento para la Utilización de la Tarjeta de Registro Bibliográfico: Manual de DOCPAL No. 2. (Procedure for using the Bibliographic Record Card: DOCPAL Manual No. 2). Santiago: CELADE (CEPAL/CELADE/L.2, Serie B No. 42).
- CEPAL. 1892. Manual de Procedimiento No. 1: Uso de Hojas de Trabajo: Hoja de Descripción Bibliográfica (HDB) y Hoja de Análisis de Contenido (HAC) (Procedure Manual No. 1: Use of the Worksheets: Bibliographic Description Worksheet (BDW) and the Content Analysis Worksheet (CAW). Santiago: CEPAL (Version in English in press)
- CICRED. 1979. Multilingual Population Thesaurus. Paris: CICRED.
- Corbin, John. 1981. Developing Computer-Based Library Systems. Phoenix: Oryx Press.
- Griffiths, José-Marie. 1981. Application of Mini-computers to Information Handling. Prepared for the General Information Programme and UNISIST. Paris: UNESCO. (PGI-81/WS/28).
- Dierickx, Harold and A. Hopkinson. 1981. Reference Manual for Machine-Readable Bibliographic Descriptions. 2nd revised edition. Paris: General Information Programme and UNISIST, UNESCO. (PGI/81/WS/22).
- Dierickx, Harold. 1982. A proposed common format for existing and projected computerized bibliographic information systems in the Caribbean, resulting from a mission in the Region from 22 November to 18 December 1981. London: UNIBID (Photocopied typescript; February, 1982).
- Habitat. 1975. Data Management for Urban and Regional Development. Nairobi: UN Center for Human Settlements (Habitat) (CHS/PP/81-1/S).
- ISO. 1977. ISO Standards Handbook 1, 1977: Information Transfer. Geneva: ISO.
- Lancaster, F. Wilford. 1968. Information Retrieval Systems: Characteristics, Testing and Evaluation. New York: John Wiley and Sons.
- OECD. 1979. OECD Macrothesaurus para el Tratamiento de la Información relativa al Desarrollo Económico y Social. Paris: OECD.
- Meadow, Charles T. and Pauline Cochrane. 1981. Basics of On-Line Searching. New York: John Wiley and Sons.
- Morin-Labatut, Risele and Maureen Sly. 1982. Manual for the preparation of records in development-information systems. Recommended Methods for Development-Information Systems, Vol. 1. Ottawa, Ontario: IDRC. (IDRC-TS40e).

- Park, M.K., and R.E. Stearns, Jr. 1982. Evaluation of the IMD IV+V system software package for UNISIST Reference Manual Implementations. Paris: UNESCO. (Photocopied typescript; May 1982).
- POPIN. 1982. Report on the First POPIN Advisory Committee Meeting. Geneva, 22-25 March 1982. New York: United Nations Population Division.
- POPLINE. 1981. POPLINE Thesaurus. Compiled by the Center for Population and Family Health of Colombia University; the Population Information Program of John Hopkins University; and Population Index of Princeton University.
- Toong, Hoo-min D., and Amar Gupta. 1982. Personal Computers. Scientific American. 247:86-107 (December 1982).
- UNESCO. 1980a. UNISIST Guide to Standards for Information Handling. Prepared by the UNISIST Working Group on Bibliographic Data Interchange. Compiled by Erik Vajda. Paris: UNESCO.
- UNESCO. 1980b. Guide for the Establishment and Evaluation of Services for Selective Dissemination of Information. Paris: UNESCO. (PGI/WS/80/14).
- United Nations. 1982. National Household Survey Capability Programme: Survey Data Processing: A Review of Issues and Procedures. Compiled by the United Nations Department of Technical Cooperation for Development and the Statistical Office. New York: United Nations. (DP/UN/INT-81-041/1 Draft).
- Valantin, Robert. 1981. CDS/ISIS and MINISIS: A Functional Analysis and Comparison. Ottawa, Ontario: IDRC (IDRC-TS37e).

TABLES AND FIGURES

Table 1. 1A: Development and database creation costs.
1B: Operational costs.

Table 2. Important computer/software (non-commercial)
combinations for bibliographic information
storage and retrieval.

Figure 1. Simplified Diagram of a computer.

Figure 2. Telecommunication links with a computer.

Figure 3. Simplified diagram of a computer record
and its fields.

Figure 4. Serial and inverted file searches.

Table 1

Table 1A Development and Database creation costs

Consultants to assist in the design of the system	Site preparation (including in- installation of cables, air con- ditioning for the computer, etc.)
Hardware	Staff training
System software	Additional staff if required
Bibliographic storage and retrieval software	Programmers, systems analysts
Applications programming	Entry of earlier materials
Other equipment (desks, etc; magnetic tapes, disks, etc)	Supplies
Travel to visit other computerized PI Units	Miscellaneous costs (Photocopying, telephone calls, etc.)

Table 1B. Operational costs

Staff salaries	Programming assistance (if not on regular staff)
Abstractors (if not on staff)	Software licenses
Computer supplies (paper, magnetic tapes, etc)	Electricity
Other supplies	Computer communications (e.g., telephone lines)
Hardware rental or leases	Miscellaneous costs (Photocopying, telephone calls, cables, etc.)
Hardware preventive mainten- ance (on equipment owed including terminals)	Printing costs (if publish a journal)
Additional equipment (e.g., additional terminals, etc)	Postage for mailing journals, etc.
	Publicity

Table 2

Computer/Software (non-commercial) Combinations
for Bibliographic Information Storage and Retrieval

Computer	Operating system	Software	Possible source
IBM Mainframe (a)	OS; CICS for on-line operation	CDS/ISIS	UNESCO, Paris
IBM Mainframe (a)	DOS, DOS/VS, or DOS/VSE	ISIS	IDRC, Ottawa
IBM Mainframe (a)	DOS-VM-CMS	ISIS	CELADE, Santiago
IBM Mainframe (a)	OS; CMS for on-line operation (under development)	ISIS	CEPAL, Santiago
Hewlett-Packard 3000 series mini-computer		MINISIS	IDRC, Ottawa
PDP 11/70 (mini-computer) or Philips 857 mini- computer (b)		DOMESTIC	National Center of Scientific and Technological Infor- mation (COSTI), Tel Aviv and KTS- Informations System GmbH, Munich
PRIME 300 mini- computer (b)	PRIMOS	STATUS II	United Kingdom Atomic Energy Authority
UNIVAC (*)		ISIS (Cobol version)	Phillipine agency

(a) IBM 360, 370, 43xx or 30xx series mainframe computers.

(b) Cited in Griffiths (1981).

(c) Cited in Valantin (1981:8).

Figure 1. Simplified Diagram of a computer.

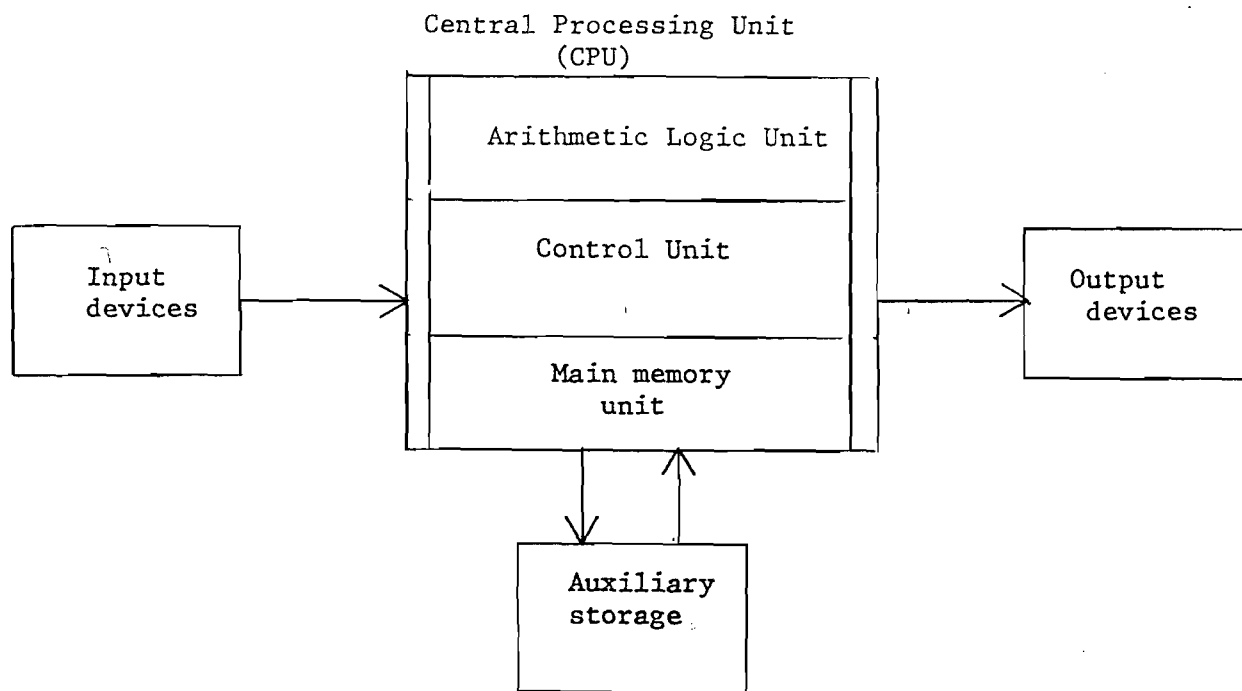


Figure 2. Telecommunication links with a computer.

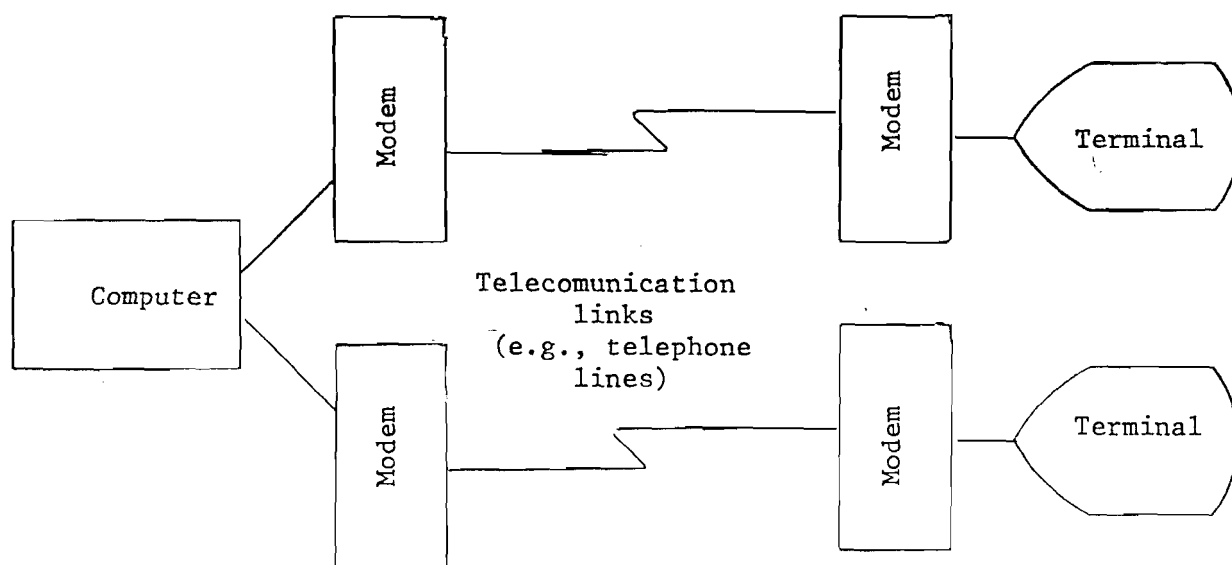


Figure 3. Simplified diagram of a computer record and its fields.

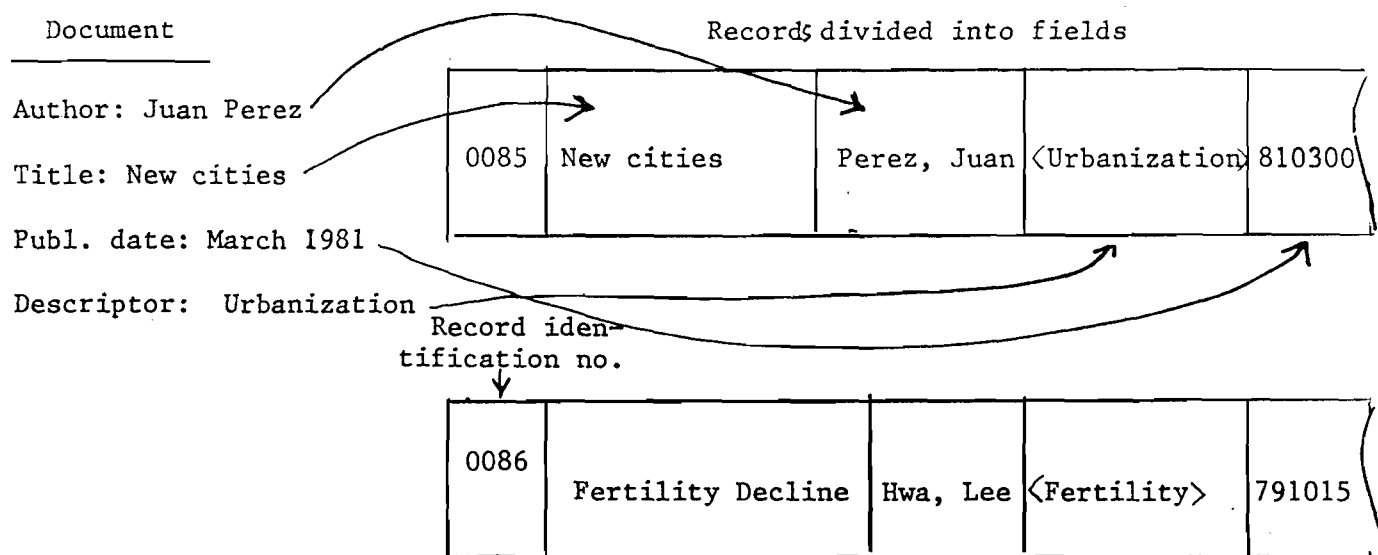
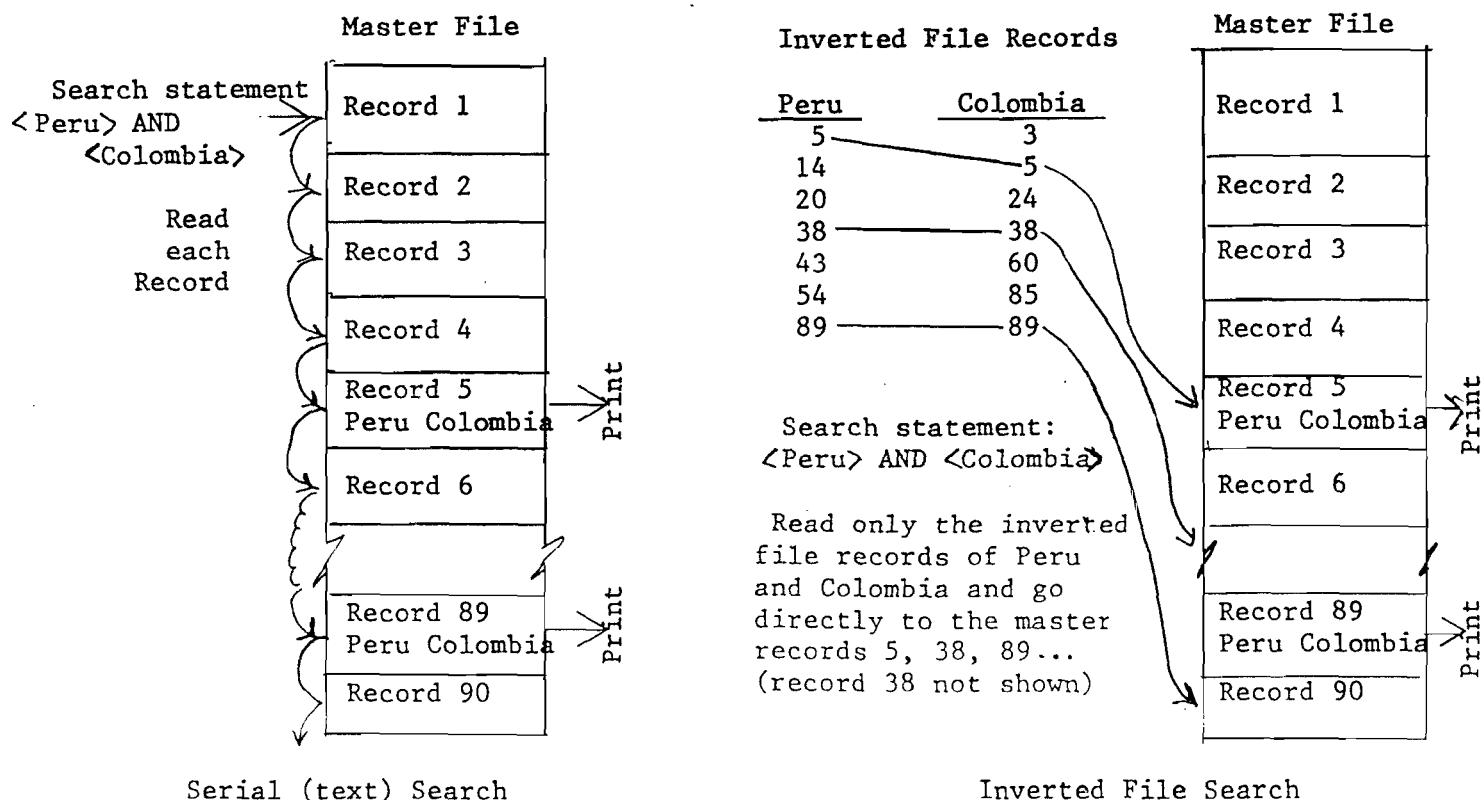


Figure 4. Serial and inverted file searches.



ANNEX 1

INPUT AND OUTPUT FLOW DIAGRAMS

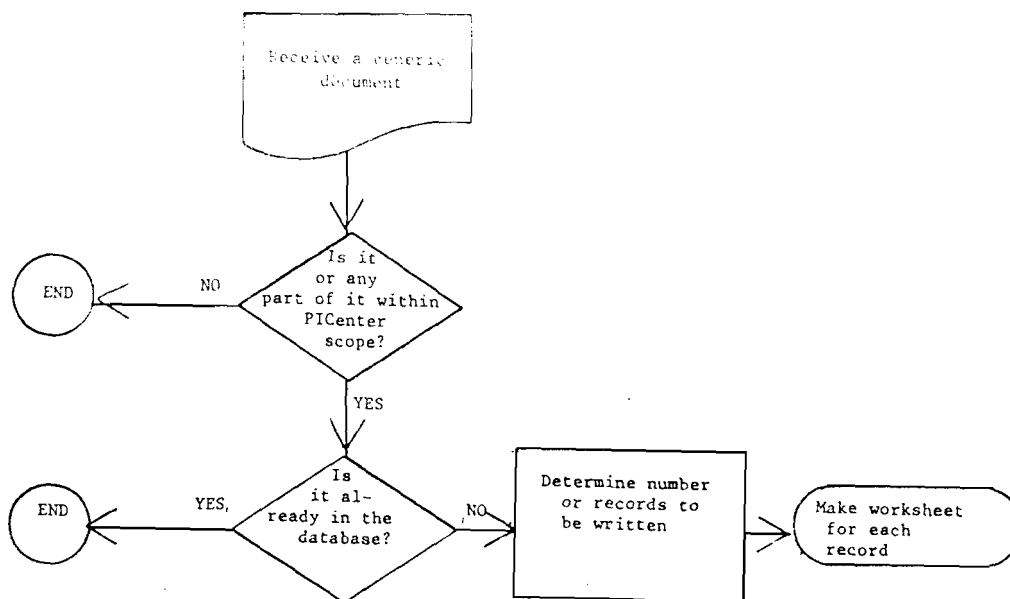
Annex 1A. Input flow diagram: Selection of documents.

Annex 1B. Input flow diagram: Processing of documentary units.

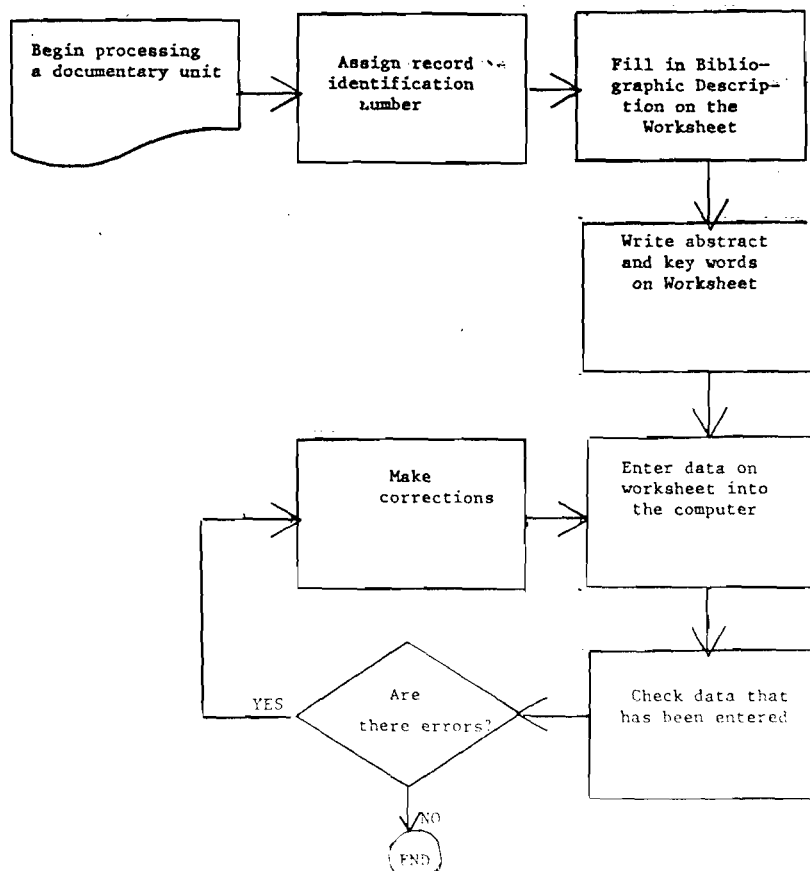
Annex 1C. Output flow diagram: Interactive searches.

Annex 1D. Output flow diagram: Production of a journal.

ANNEX 1A: INPUT FLOW DIAGRAM: SELECTION OF DOCUMENTS

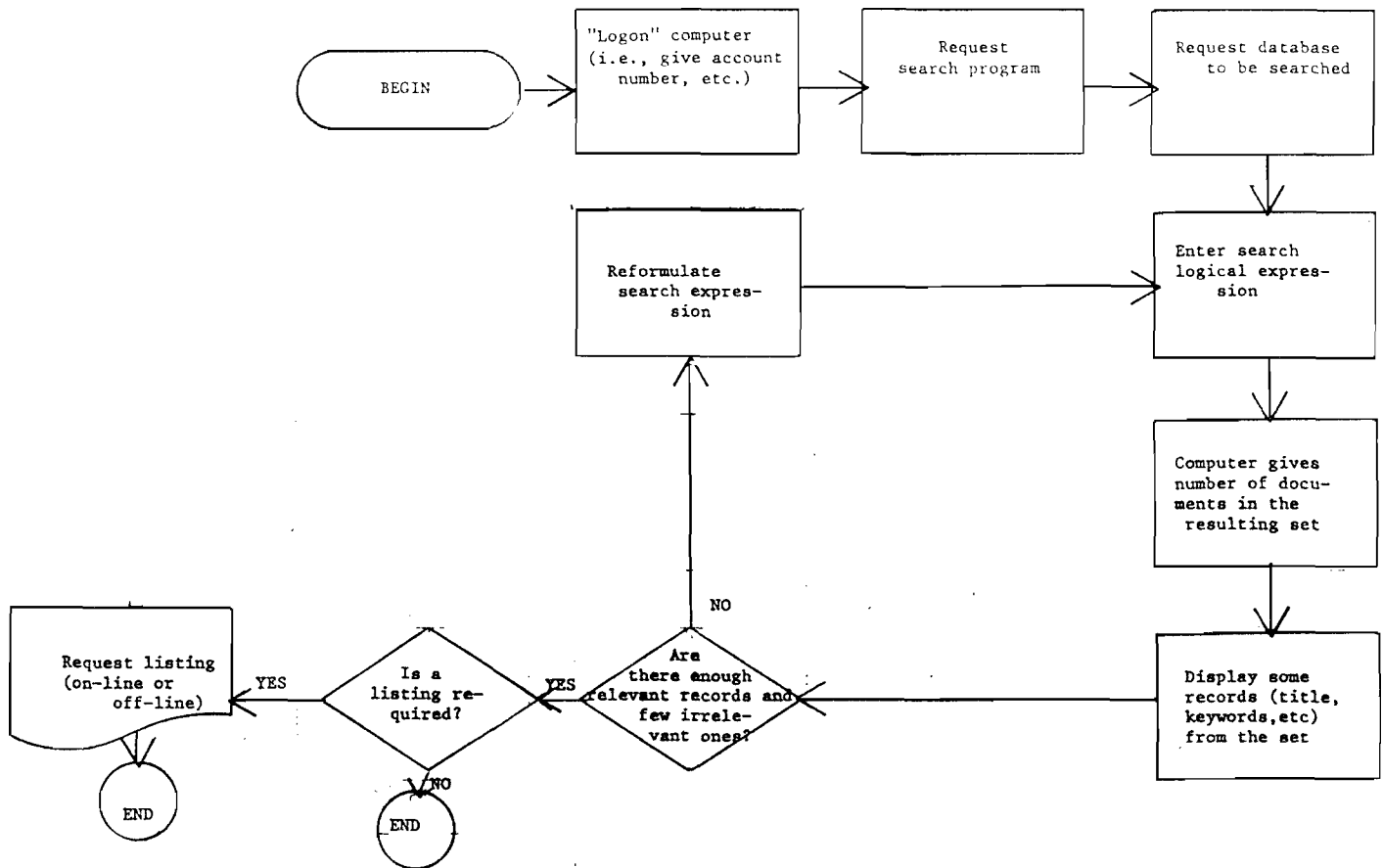


ANNEX 1B: INPUT FLOW DIAGRAM: PROCESSING OF DOCUMENTARY UNITS

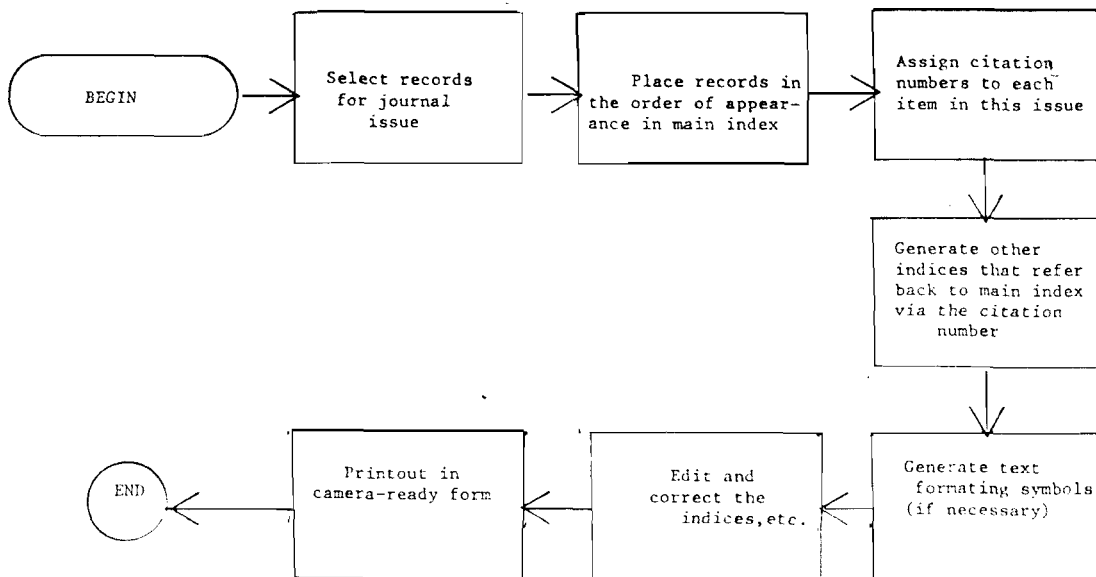


In some software systems (preferred), the information from a newly entered record is automatically added both to the Master File and the inverted file. In others, it is necessary to merge the new record(s) into the Master File and then carry out an inversion procedure with the entire database to create a new inverted file.

ANNEX 1C: OUTPUT FLOW DIAGRAM: INTERACTIVE SEARCHES



ANNEX 1D: OUTPUT FLOW DIAGRAM: PRODUCTION OF A JOURNAL



ANNEX 2: SEARCH LOGIC

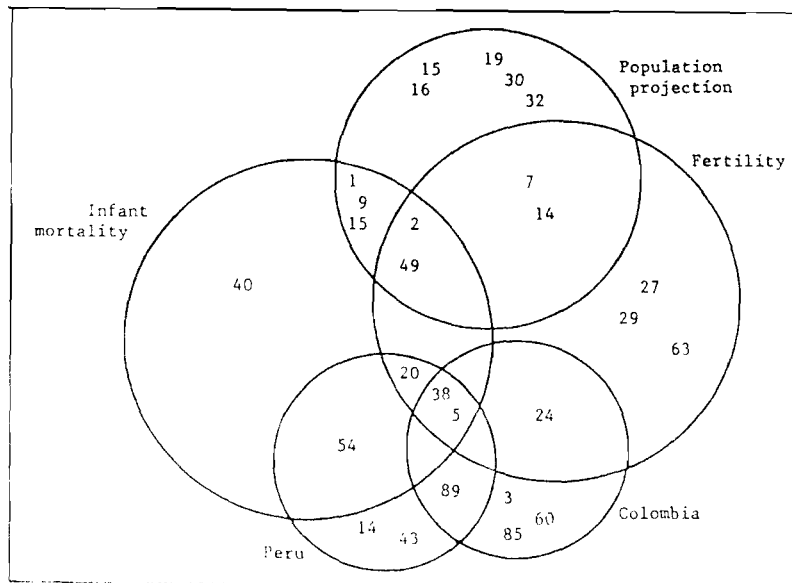
A search for bibliographic information normally has the dual objectives of (a). identifying all the records representing documents in the database that meet the user's request (high recall), while (b). excluding all unwanted records (high precision). See Lancaster, 1968:54-63 for more on recall and precision.

Before beginning a search, the concepts in the user's request must be translated into the terms employed in the database; these may be descriptors, countries or geographical areas, author names, dates, etc. During the search the terms are combined together in logical expressions using Boolean Algebra; this involves the combination of sets through logical operations like AND, OR, etc. The items in the sets are records; for example, the set <Fertility> comprises the records in the database that have the descriptor "Fertility".

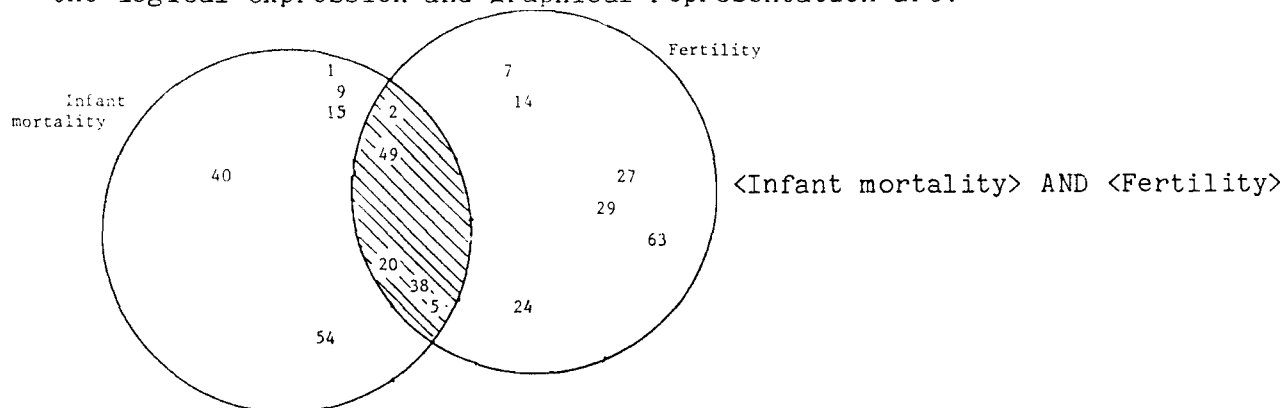
To demonstrate the use of logical expressions in a search, consider the sets of records in a database for each of the terms below; the numbers are record identification numbers, although for convenience the leading zeros have been eliminated, i.e., 0089 is shown as 89. Each of these sets could be on a computerized inverted file record (Section 3B) or a manual UNITERM card (Section 8).

<Infant Mortality>	<Fertility>	<Population Projection>	<Peru>	<Colombia>
-----	-----	-----	-----	-----
1	2	1	5	3
2	5	2	14	5
5	7	7	20	24
9	14	9	38	38
15	20	14	43	60
20	24	15	54	85
38	27	16	89	89
40	29	19		
49	38	30		
54	49	32		
	63	49		

Graphically the entire database can be represented by a rectangle and the sets defined, which may have other records as well, by circles:

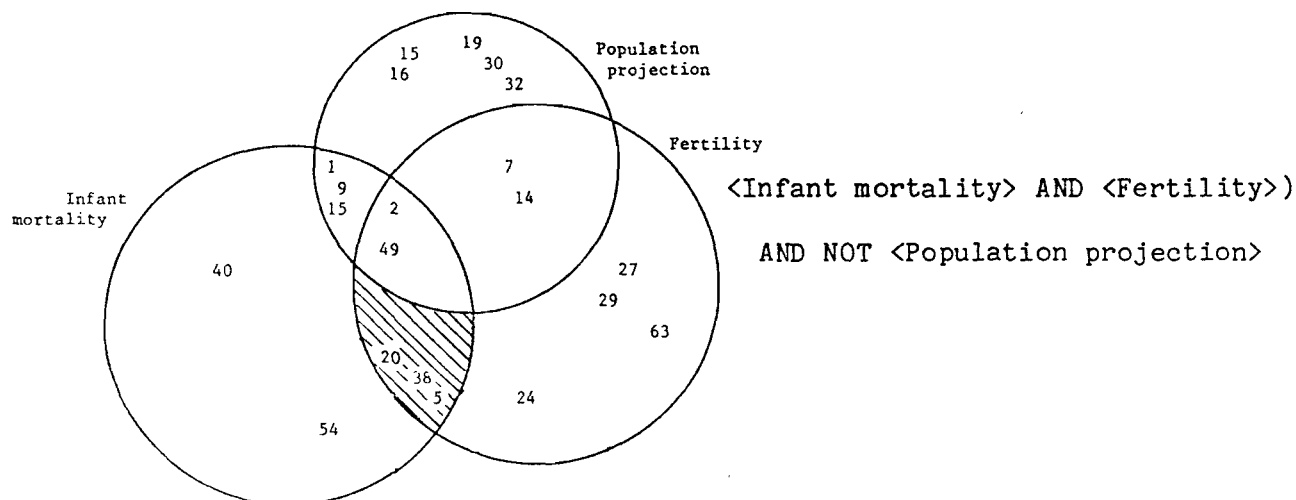


If a user requests all records treating both infant mortality and fertility, the logical expression and graphical representation are:



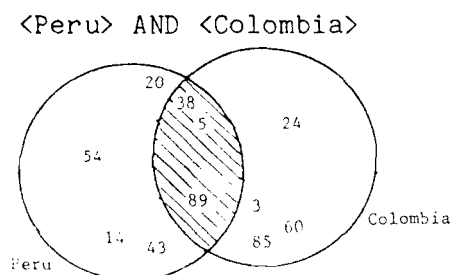
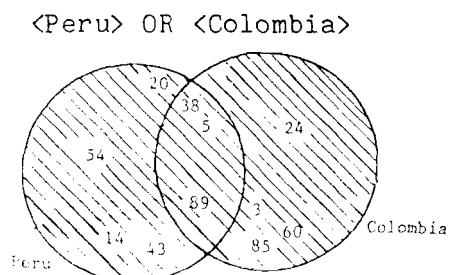
The result of the logical operation AND, often called an intersection, is the new set shown with shading, which has the records 2, 49, 20, 38, and 5.

If the user wants records considering only the effect of infant mortality on fertility, the searcher should eliminate records on population projections since these treat infant mortality and fertility as projection inputs. The logical operation AND NOT will exclude the unwanted records:



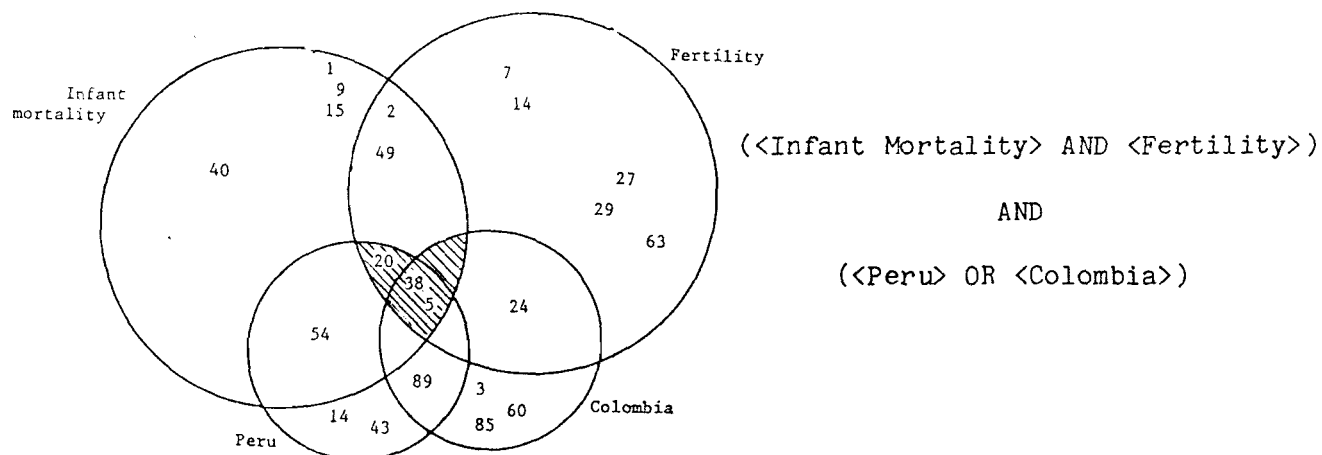
The parenthesis indicates that the new set (<Infant mortality> AND <Fertility>) is formed before proceeding to the next operation.

The result of the logical operation OR will locate all records that treat, for example, Peru or Colombia, or both. This operation, sometimes called a union, produces a different set from that generated by the logical operation AND, as can be seen by comparing the sizes of the shaded areas:

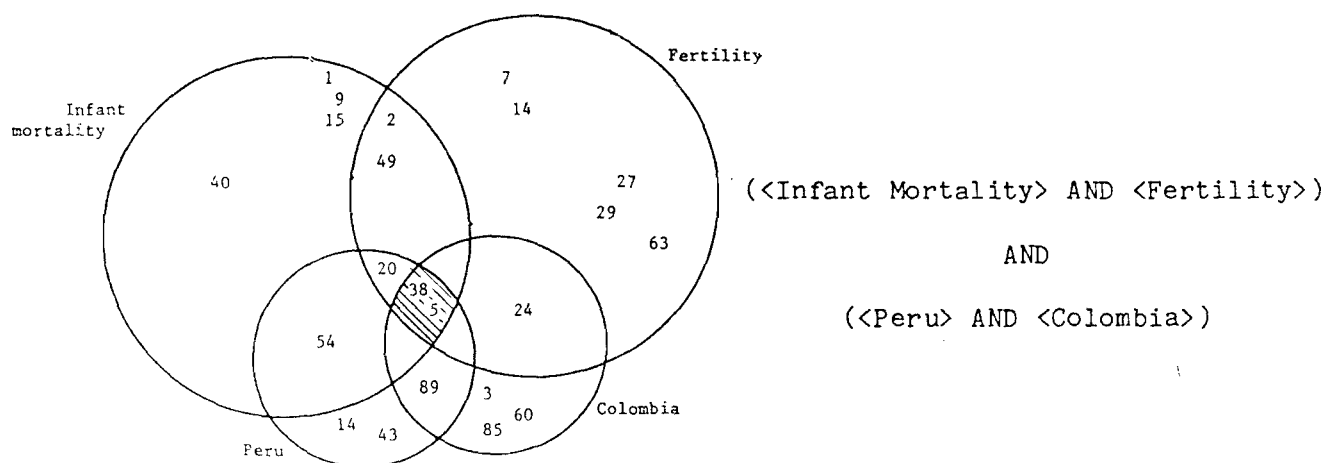


Another related logical operation, less frequently used, is the "exclusive or" written XOR, which would include documents treating Peru or Colombia, but not both, i.e., only records 20, 54, 14, 43, 24, 3, 60 and 85 would remain and 38, 5 and 89 would be excluded.

Once a new set is generated from a logical operation, it can be combined in turn with other sets. For example, a request for records treating infant mortality and fertility in Peru or Colombia, would be:



However, if only comparative studies of Peru and Colombia are wanted, the result would be:



ANNEX 3

WORKSHEETS AND WORKCARDS

Annex 3A. Worksheet used by CEPAL and CELADE/DOCPAL.

Annex 3B. Worksheet recommended by IDRC.

Annex 3C. Manual system workcard and UNITERM card compatible
with the CELADE/DOCPAL computerized system.

Annex 3D. Manual system worksheet used by RIPS/PIDSA.



CIPAL BIBLIOGRAPHIC INFORMATION SYSTEM

B D W 1

[illegible]

Form. BDW	Bibl. Inform. Syst.	CEPAL	October 1982	82-9-2092
-----------	---------------------	-------	--------------	-----------

B D W 2

I M P R I N T	Publisher	38			
	City	39		Country	40
	Edition	41		Descriptive information	42
	Date of publication	43		Standard date of publication	44
	Symbol	45			
A R E A	R				
	ISBN	47		Sales No (UN)	48
T H E S I S	Institution	50			
	R				
	Degree	51			
C O N F E R E N C E	Sponsoring organization	52			
	R				
	Name	53			
	City	54		Country	55
	Date of conference	56		Standard date of conference	57
P R O J E C T	Sponsoring organization	58			
	R				
	Name	59			
	R			Number	60
C O M P L E M E N T A R Y	Dissemination	62		Type of document	63
	Language of text	64		Language of abstract	65
	Other language editions	66	Document in other language:		
	Related documents	67	Document in other source:		
	Notes	68	Complementary document:		
I N F					
Fields to copy		69	Copies		
			ISIS No		
C O N T R O L	Dates	90	BDW	Documentable	91
	R				
	State of record	92			
Comments					

ANNEX 3A: Worksheet used by CEPAL and CELADE/DOCPAL (first sheet, front and back)



UNITED
NATIONS

CONTENT ANALYSIS WORKSHEET
CEPAL BIBLIOGRAPHIC INFORMATION SYSTEM

CAW 1

☐ New ☐ Correction

ISIS No

Name of file	01	Author and title	Pages to abstract
Accession No	02		

Abstract for sales catalogue	71
------------------------------	----

Abstract	72
----------	----

Number of references	73
----------------------	----

Period covered			
Initial year	74	Final year	75

Form CAW B/N. Inform. Syst. CEPAL October 1982 42-9-2092

CAW 2

DESCRIPTOR	Subject analysis	76	
	Statistical data	77	
CATEGORIES	Primary	80	
	Secondary	81	
	Geographic	82	
COUNTRIES	Primary	83	AG AN AR BB BO BR BS BZ CL CO CR CU DM DO EC FK GD GF GP GT GY HN HT JM KN KY LC MQ MS MX NI PA PE PR PY PZ SR SV TC TT UY VC VE VG XC XI XL XS
	Secondary	84	XZ ZZ
	Administrative divisions and natural regions	85	
Acronyms	86		
Relationship with other systems	87		
CONTROLS	CAW 1	R 90	CAW 1 -
	CAW 2	R 90	CAW 2 -
	Participating centre sent	R 92	1 -
	Coordinating centre received	R 92	2 -
Comments	Documentalist	R 91	CAW 1 -
	Documentalist	R 91	CAW 2 -
	Final revision	R 92	3 -

ANNEX 3A (continued): Worksheet used by CEPAL and CELADE/DOCPAL
(second sheet, front and back)

ISN																																																																
PCN A110																																																																
LIBCAL ★	LIBCOD A121																																																															
	CALLNO A122																																																															
<table border="1"> <tr> <td colspan="10">RELATED RECORD ★</td> </tr> <tr> <td>RELATR A141</td> <td>RELRN A142</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>RELATR A141</td> <td>RELRN A142</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>RELPCN A143</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>RELPCN A143</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table>											RELATED RECORD ★										RELATR A141	RELRN A142										RELATR A141	RELRN A142										RELPCN A143											RELPCN A143										
RELATED RECORD ★																																																																
RELATR A141	RELRN A142																																																															
RELATR A141	RELRN A142																																																															
RELPCN A143																																																																
RELPCN A143																																																																
<table border="1"> <tr> <td>AGRI A180</td> <td>LITERARY INDICATORS</td> <td>CONFERENCE K</td> <td>DICTIONARY L</td> <td>NUMERICAL DATA N</td> <td>THESES OR DISSERTATION U</td> <td>LEGISLATION W</td> <td>BIBLIOGRAPHY Z</td> <td>MAPS INCLUDED Y</td> <td>SUMMARY E</td> <td>NONCONVENTIONAL V</td> <td>REVIEW R</td> </tr> </table>											AGRI A180	LITERARY INDICATORS	CONFERENCE K	DICTIONARY L	NUMERICAL DATA N	THESES OR DISSERTATION U	LEGISLATION W	BIBLIOGRAPHY Z	MAPS INCLUDED Y	SUMMARY E	NONCONVENTIONAL V	REVIEW R																																										
AGRI A180	LITERARY INDICATORS	CONFERENCE K	DICTIONARY L	NUMERICAL DATA N	THESES OR DISSERTATION U	LEGISLATION W	BIBLIOGRAPHY Z	MAPS INCLUDED Y	SUMMARY E	NONCONVENTIONAL V	REVIEW R																																																					
STATUS A150	CD																																																															
ORIGIN A160																																																																
AGRI RECTYP A170																																																																
ACCESS A190																																																																

LEVDU A131	AMC	LEVGD A132	MCS
---------------	-----	---------------	-----

DOCUMENTARY UNIT (A/M/C)	PERSONAL AUTHOR ★	B210		
	INSTITUTION ★	B220		
	TITLE OF DOCUMENTARY UNIT ★	B230		
	TRANSLATED TITLE ENGLISH	B240		
	TRANSLATED TITLE			

GENERIC DOCUMENT (M/C/S)	PERSONAL AUTHOR (M/C) ★	B310		
	INSTITUTION (M/C) ★	B320		
	TITLE OF GENERIC DOCUMENT (M/C)	B330	VOLUME/PART (C) B340	
	TITLE OF SERIAL (S)	B410		
	VOL/PART (S)	B420	ISSN (S)	B430

LANGUAGES	TEXT B540			SUMMARIES B560		
PUBLISHER	NAME B611					
	CITY B612					
DATE OF IMPRINT/ISSUE	FREE FORM B620				YEAR B630	
COLLATION	PAGINATION DESCRIPTION B641			NOTES B642		
REPORT NUMBER OR SYMBOL ★	B650					
ISBN	B660			EDITION B670		
MEETING, CONFERENCE, ETC. ★	NAME B711					
	CITY B712	COUNTRY CODE B713		DATE B714		
	NAME B711					
	CITY B712	COUNTRY CODE B713		DATE B714		
ACADEMIC DEGREE	B720					
PATENT	IPC B731			NUMBER B732	COUNTRY CODE B733	
DEVELOPMENT PROJECT ★	NAME B741					
	NUMBER B742			VALUE B743		
AVAILABILITY	B810					
ANCILLARY DATA ★	B820					
PQEO C100 ★				BQEO C110		AGRCAT C120
DEVPUR C130		MAINLO C140		DEVSEC C150		SECLO C160
CORRYS C170 ★						ANNEX
DESCRIPTORS	C210					
	C220					
ADMINISTRATIVE DIVISION ★	C410					
NATURAL REGIONS	C430					
COMPLETED BY	D100	CHECKED			INPUT/CHANGE D120/D140	

ANNEX 3B: Worksheet recommended by IDRC (first sheet, front and back)

ANNEX

PCN		-		-					
ISN									

ABSTRACT C310

STATISTICAL TABLE ★ C510

OTHER; PLEASE SPECIFY

(/continued) Worksheet recommended by IDRC (second sheet)

rb

Parietas Unitermino	04	06	13	16	47	50	26	29	36	41	87	88	78	Total =
---------------------	----	----	----	----	----	----	----	----	----	----	----	----	----	---------

File: CountriesFile: Countries

Annex 3D. Manual system worksheet used by RIPS/PIDSA.

(TO BE OBTAINED)

ANNEX 4

EXTERNAL COMPUTERIZED SEARCH SERVICES

Annex 4A. POPLINE Literature Search Request Form
and an Example of a Search Listing.

Annex 4B. CELADE/DOCPAL Literature Search Request
and Document Copy Request Forms



POPLINE, a computerized literature search service, offers information on all aspects of family planning and population, including human fertility, contraception, family planning programs, evaluation, and related health, law, and policy issues. Updated monthly, POPLINE contains bibliographic citations and abstracts in English of over 75,000 documents, published and unpublished. Over half of the documents in the system are from non-US sources and 10 percent are in languages other than English.

POPLINE is produced by the Population Information Program of the Johns Hopkins University and the Library/Information Program of the Center for Population and Family Health of Columbia University in cooperation with the US National Library of Medicine

and with financial support from the United States Agency for International Development. POPLINE services include retrospective searches, current awareness searches, and document delivery.

Availability

Developing Countries: Searches can be requested directly from one of the organizations listed below. All searches are free of charge.

United States: Available through the National Library of Medicine's MEDLARS system, POPLINE can be accessed at over 1,200 academic, commercial, and government organizations that are designated search centers. Contact your librarian or the National Library of Medicine, Bethesda, Maryland 20209, for the center nearest you.

Other Developed Countries: Searches can be requested directly from one of the organizations listed below. Literature searches for persons employed by international development and training institutions are free of charge. A fee may be charged to all other requesters.

POPLINE Literature Search Request

A request for a POPLINE literature search, submitted on this request form or in similar format, will be filled promptly.

Please define your subject as specifically as possible (for example, the prevalence of oral contraceptive use in Thailand since 1973). Include, where appropriate, the purpose of the search (for example, a scholarly article on your own research, a lecture to a class, or an advisory report to an international agency), dates of the documents to be covered (for example, documents produced in 1978-1980), and geographic limitation (for example, Latin America only).

SEARCH TOPIC: _____

PURPOSE: _____

ORGANIZATION: _____

Mail this request for your POPLINE search to either of these addresses:

Population Information Program

Attn: POPLINE
The Johns Hopkins University
624 North Broadway
Baltimore, Maryland 21205
USA

Center for Population and Family Health

Attn: POPLINE
Columbia University
60 Haven Avenue
New York, New York 10032
USA

(Print or type all information clearly)

NAME: _____

ADDRESS: _____

CITY: _____ POSTAL CODE: _____

COUNTRY: _____

Example of POPLINE search listing

00015817

LATIN AMERICA (SDI)

PAGE 29

- 49 SI - IND/0020522
 AU - Reichert JS
 TI - The migrant syndrome: seasonal U.S. wage labor and rural development in Central Mexico
 GN - PI vol. 47, accession no. 3614
 AB - The author analyzes the effects of seasonal migration on the economic development of the town of Guadalupe, Mexico. He attempts to identify the differences in earning of families with legal, illegal, and no migrants and compares patterns of consumption between the three family groups. He also discusses the implications of U.S. labor migration for the sending communities of rural Mexico
 SO - Human Organization. 1981;40(1):56-66.
- 50 SI - CPFH/11961cr982
 AU - Rosenzweig MR ; Schultz TP
 TI - Child mortality and fertility in Colombia: individual and community effects.
 AB - The education of a mother is strongly and positively correlated with the survival rate of her children. This paper combines household data from the Colombian Census of 1973 and characteristics of the 900 residential areas in Colombia, to test various hypotheses concerning the mechanism by which mother's education and public policies affect child survival and the distribution of health benefits resulting from policy interventions. The hypothesis is advanced that education provides people with skills in acquiring and decoding new information and thus effectively lowers the costs of using more beneficial child health and contraceptive technologies. Since a primary function of health and family planning programs is to disseminate information on these same technologies, the hypothesis is tested that mother's education and these program interventions may substitute for each other in improving child health and reducing family size. The empirical analysis confirms that in urban areas the availability of medical services, family planning activities, transportation infrastructure and climate, in addition to mother's education, are associated with child mortality ratios and fertility within a birth cohort of mothers. The least educated mothers are the most strongly affected, in terms of their reduced fertility and increased child survival rates, by the local urban health programs. The evidence is thus consistent with the substitution hypothesis. No effect of program interventions and medical facilities are found on rural populations, though both child mortality ratios and fertility are lower for more educated rural women. (author's)
 AD - Dept. of Economics, Univ. of Minnesota, 1035 Business Administration Building, 271 19th Avenue, Minneapolis, MN 55455

CELADE/DOCPAL Literature Search Form



DOCPAL / CELADE

SOLICITUD DE BUSQUEDA POR COMPUTADOR
COMPUTERIZED SEARCH REQUEST

BUS

Fecha rec./Date rec.

Fecha env./Date sent

Instrucciones al reverso/See instructions on other side

Nombre/Name		Institución/Institution	
Dirección/Address			
1. Tema de la búsqueda (sea específico) / Search topic (be specific):			
2. Cobertura geográfica / Geographic coverage:			
<input type="checkbox"/> Cualquier país de América Latina o El Caribe/Any countries of Latin America or the Caribbean <input type="checkbox"/> Solo (especifique)/Only (specify):			
3. Idiomas de los documentos / Languages of the documents:			
<input type="checkbox"/> Inglés/English <input type="checkbox"/> Portugués/Portuguese <input type="checkbox"/> Francés/French <input type="checkbox"/> Español/Spanish <input type="checkbox"/> Otro (especifique)/Other (specify):			
4. Instrucciones especiales / Special instructions			
Fecha búsqueda/Date search	por/by	Comentarios/Comments	Fecha env./Date sent

Form. DOCPAL Solbus 15/12/77

CENTRO LATINOAMERICANO DE DEMOGRAFIA
SISTEMA DE DOCUMENTACION SOBRE POBLACION EN AMERICA LATINA

SOLICITUD DE BUSQUEDA POR COMPUTADOR

Las bibliografías especializadas hechas mediante búsquedas computarizadas en la base de datos DOCPAL es un servicio que se entrega sin costo alguno para el solicitante. Esta base de datos incluye documentos tanto publicados como inéditos (libros, capítulos de libros, artículos, trabajos presentados a conferencias, informes, etc.) relacionados con temas de población producidos a partir de 1970 en América Latina y El Caribe o sobre la región. A Diciembre de 1977 el Sistema contaba con 4000 documentos a los que se agregan mensualmente alrededor de 250 items.

Instrucciones

1. Llene el formulario al reverso. Cada consulta debe presentarse en un formulario aparte; si no se cuenta con un número suficiente de formularios, se puede usar papel corriente.
2. La consulta debe presentarse en la forma más específica posible. Por ejemplo, "el efecto de la reforma agraria o políticas agrícolas sobre la migración rural a áreas metropolitanas", o "todos los documentos escritos por Jorge Somoza sobre mortalidad a partir de 1974".
3. Incluya su dirección completa a fin de que el correo funcione en forma expedita. Serán enviadas por vía aérea.
4. Envíe sus solicitudes a Búsquedas DOCPAL, CELADE, Casilla 91, Santiago, Chile.

COMPUTERIZED SEARCH REQUEST

The preparation of specialized bibliographies based on computerized searches of the DOCPAL database is a service which is provided free of charge to those requesting it. The database includes published and unpublished documents (books, chapters of books, articles, conference papers, reports, etc.) concerning population, written since 1970 in Latin America and the Caribbean or about the Region. In December 1977 there were 4,000 documents in the System and over 250 are added each month.

Instructions

1. Fill in the form on the other side. Each individual search should be placed on a separate form; ordinary paper may be used if additional forms are not available.
2. State the topic as specifically as possible. For example, "the effect of agrarian reform or agricultural policies on rural-urban migration to capital cities", or "all documents written by Jorge Somoza on mortality since 1974".
3. State your full address to insure rapid delivery. Results will be sent airmail.
4. Send your requests to Búsquedas DOCPAL, CELADE, Casilla 91, Santiago, Chile.

Form. DOCPAL Solbus 15/12/77

Document Copy Request Form

DOCPAL / CELADE

SED

--	--

N° de Factura/Invoice N°

SOLICITUD DE ENTREGA DE DOCUMENTOS
DOCUMENT REQUEST FORM

Favor leer instrucciones y firmar al reverso/Please read instructions and sign on the other side

Nombre/Name
Institución/Institution
Dirección, Ciudad, País/Address, City, Country

NO ESCRIBIR EN AREAS SOMBRADAS / DO NOT WRITE IN SHADED AREAS

Form. DOCPAL Solentreg 15/12/77

Pág. fotocop. Pag. copied	Comentarios Comments	N° Resumen Abstract N°	N° Archivo DOCPAL DOCPAL File N°	Apellido 1 ^{er} Autor Surname 1 st Author	Instrucciones Especiales Special Instructions

Pág. fotocop. Pag. copied	Costo/Cost		PAGAR A CELADE	PAY TO CELADE	Intercamb. fotocop. Photocop. exchange	Fecha Date	Copiado/Copied	Enviado/Sent
	Fotocop. Photocopies	Publicaciones Publications	US\$		pag.		CELADE Publ.	Pago rec./Pay rec.

NSC = No se puede copiar/Cannot be copied

NO DOC. = Documento no en DOCPAL/Doc. not in DOCPAL

LE = Libro entero/Whole book

R = Restring./Restricted

CENTRO LATINOAMERICANO DE DEMOGRAFIA
SISTEMA DE DOCUMENTACION SOBRE POBLACION EN AMERICA LATINA

SOLICITUD DE ENTREGA DE DOCUMENTOS

DOCUMENT REQUEST FORM

INSTRUCCIONES

1. Complete ambos lados del presente formulario. La declaración que usted debe firmar, explica las condiciones y costos. Se entrega información adicional en la Sección "Servicio de Entrega de Documentos" de la Revista DOCPAL Resúmenes sobre población en América Latina.
2. Cite la referencia completa, si desconoce el Número de Resumen o del Archivo DOCPAL. No se entregan copias de libros enteros.
3. Envíelo a DOCPAL, CELADE, Casilla 91, Santiago, Chile.

INSTRUCTIONS

1. Fill in both sides of this form. The declaration, which must be signed, explains the terms and costs. Additional information is given in the "Regional Clearinghouse" Section of the Journal DOCPAL Latin American Population Abstracts.
2. If the Abstract and/or DOCPAL File Number is not known, cite the complete reference. Entire books cannot be copied.
3. Return to DOCPAL, CELADE, Casilla 91, Santiago, Chile.

DECLARACIÓN

Declaro que las copias aquí solicitadas son solamente para propósitos de estudio e investigación y reemplazan un préstamo del original o su transcripción manual. Las copias no serán reproducidas o usadas con fines comerciales.

Entiendo que si resido en América Latina o El Caribe, puedo solicitar gratuitamente un número razonable de documentos. Si resido fuera de la región, acepto pagar US\$ 0,15 por página fotocopada contra factura enviada por CELADE.

DECLARATION

I hereby state that the copies requested here are for research or study purposes only and in place of a loan of the original or its manual transcription. The copies will not be reproduced or used for commercial purposes.

I understand that if I live in Latin America or the Caribbean, I can request a reasonable number of documents free. If I live in another region, I agree to pay US\$ 0,15 per photocopied page on receipt of the bill from CELADE.

Firma/Signature

Fecha/Date

FIRMAR Y LLENAR AL REVERSO / SIGN AND FILL IN OTHER SIDE

INDEX

PAGES ON WHICH TERMS ARE DEFINED OR USED SUBSTANTIVELY

(The numbers refer to the pages on which the terms are underlined
and their definitions and/or utilization are explained.)

AGRIS, 16
Alphanumeric character, 3
Analytic level, 13
AND, 36
AND NOT, 36
Application program, 4
Arithmetic-logic unit, 3
Automatic inversion, 17
Auxiliary storage, 3, 13, 16, 21

Back-up computer, 11, 21
Back-up copies, 17
Batch mode, 4, 17
Batch search, 18
Bibliographic database, 7
Bibliographic description, 14
Bibliographic information, 6
Bibliographic level, 9, 13, 14
Bibliographic material, 2
Bibliographic record, 6
Bit, 3
Boolean Algebra, 7, 35
Bug, 4
Byte, 3

Camera-ready output, 21
Card, 6
Carrier format, 14
Carrier language, 21, 23
CDS/ISIS, 20
CELADE/DOCPAL, 15, 22, 23
Central processor unit, 3
CEPAL, 15
Character set, 21
Chips, 3
Codes (countries, language), 16
Collective level, 13
Combination of sets, 35
Compatibility, 13, 15
Compatible input formats, 13
Compiled program, 4
Compiler, 4
Computer configuration, 3
Computer environment, 10
Computer hardware, 3, 18
Computer program, 4
Computer software, 4

Computer system software, 4
Computer word , 3
Control unit, 3
Controlled vocabulary, 6
CPU, 3
Crash, 8

Data element, 14, 15
Data element definition, 14
Data element set, 14, 15
Data entry, 17, 21
Data field, 6
Database, 7
Database security, 17
Deck, 6
Descriptors, 6, 16
DIALOGUE, 23
Digital computer, 3
Document, 2
Documentary unit, 9
Down-time, 8

Essential fields, 16

Field, 6
Field Definition Table, 15
Field number, 6, 15
Fields, 22
File, 6
Financial resources, 11
Flow diagram, 12
Format, 14
Fugitive literature, 13

Generalized program package, 4
Global data correction, 17

Hardcopy output, 21
Hardware, 3
Hardware maintenance, 11, 19, 20, 21
High-level programming languages, 4
Human resources, 11

IBM cards, 4, 6
IDRC, 15
IMD IV+V, 20
Indexing terms, 14
Input, 12
Input device, 3
Input processing, 12
Interactive mode, 4, 17
Interactive search, 18
Internal format, 15, 23
International communication format, 14

International exchange format, 14
Interpreter, 4
Intersection, 36
Inverted file, 7, 17, 22
Inverted file search, 7
ISIS, 4, 17
ISO standards, 16

K, 3
Keywords, 6
Kilobyte, 3
Kilobytes, 21

Local information, 14
Logical errors, 17
Logical expression, 35
Logical operation, 35, 36

Machine language, 4
Macrothesaurus, 16
Magnetic disk, 3
Magnetic tape, 3
Main file, 6
Main memory unit, 3
Main record, 6
Mainframe computer, 5, 19
Mandatory fields, 16
Manual computer-oriented system, 22
Master file, 6, 17, 22
Master record, 6, 22
Micro-computer, 5
Micro-dataset, 2
Micro-processor, 5
Mini-computer, 5, 19
MINISIS, 20
Modem, 5
Monographic level, 13

Object deck, 4
On-line, 4
Operating system, 4
Operating system maintenance, 11
Optical coincidence device, 22
OR, 36
Output, 12
Output device, 3

PASCAL, 20
Peripheral device, 3
Peripheral equipment, 3
PI Center, 1
PIDSA, 23
PMT, 16
POPLINE, 23

POPLINE Thesaurus, 16
POPULATION BIBLIOGRAPHY, 23
Population Information Center, 1
Portable software, 19, 20
Precision, 35
Preventive maintenance, 10, 21
Programming assistance, 11

Recall, 35
Record, 6
Record format, 14
Record identification number, 6, 17, 22
Record size, 16, 21
Reference Manual format, 14
Response time, 5
RM, 13

Scope definition, 9, 12
Secondary information, 6
Serial level, 13
Serial search, 6
Sets, 7
Signal transmission quality, 11
Software, 4
Software maintenance, 4, 20
Source deck, 4, 20
STAIRS, 10
Statistical data, 2
Storage and retrieval software, 18
Systems analyst, 11, 12

Tabular data, 2
Tag, 6
Telecommunication, 5
Term, 6, 35
Text formatting software, 18
Text search, 6
Thesaurus, 6
Time-sharing, 5
Turn-key system, 11
Type of bibliographic entity, 13, 14

UNIMARC format, 14
Union, 36
UNISIST Reference Manual format, 14, 15, 23
Unit of retrieval, 2
UNITERM, 22
User-friendly, 18, 20, 21

Video terminal, 21
Visual display unit, 21

Work card, 22
Worksheets, 15

COMMENTS AND SUGGESTIONS OF ARTHUR CONNING
on the
SUGGESTIONS AND COMMENTS OF POPIN ADVISORY COMMITTEE MEMBERS ET AL

on the

GUIDE TO THE
AUTOMATION OF POPULATION DOCUMENTATION
STORAGE AND RETRIEVAL SYSTEMS
(Pop Div version 12 January 1983)

Note: Pages numbers refer to the handwritten number in the text received with the 29 April letter from H.Chiang. The first paragraph on each page is taken as that at the top even if it is a continuation from the previous page. Suggested changes or additions to the text are indented. Suggestions of those who read the Guide are cited by their name and date of the letter.

I. General comments

(Specific suggestions by page number are given in Section II)

-- Latin American bias (Pasquariella 15 March, para. 2): I tried to cite operational computerized population documentation systems in the developing countries and did not realize that ESCAP had already computerized. The information on the EBIS system of ESCAP provided by Helen Kolbe should rectify this fault.

-- Microcomputers and their software (Pasquariella 15 March, para. 2; Kolbe 19 April, p.3, para. 1; Piotrow, 8 March): The revised text (see my specific suggestions below) now discusses low-cost microcomputers and associated database software. IBM ISIS and Hewlett-Packard MINISIS are treated as free non-commercial systems for manipulating large databases. It should be clear from the text that the microcomputer approach is the solution for most small Population Information Centers.

Unless there is some UN rule to the contrary, I think most publications write microcomputer as one word (or hyphenated), but not as two words.

-- Cost figures for putting a system into operation (Harris, 17 March, para. 3; Villanueva, 18 March, point 8): I agree this would be extremely valuable for a manager about to automate, but I cannot find a way to give real figures that would be useful. Both the magnitude and the distribution of costs differ radically among countries and even over short periods within a country (DOCPAL has lost most of its staff due to major increases in all costs but particularly in staff costs).

It seems best to leave the detailed list of possible costs already included, so that a manager can estimate a budget based on local conditions. Therefore, I HAVE NOT SUGGESTED ANY CHANGES.

-- Donors agencies willing to assist automation (Harris, 17 March, para. 5): This topic is outside the scope of the Guide and therefore is NOT INCLUDED. Perhaps there should be a separate POPIN publication on agencies willing to fund

population documentation projects. Those willing to do so, will surely consider automation projects.

-- Glossary of technical terms (Villanueva, 18 March, points 1 and 2): The page limitations on the Guide forced the text to be very brief on some technical issues.

To save the many pages that a glossary would occupy, I think it is best to include only the final Index which refers back to where the terms, underlined for easy location on the page, are defined or used in a context that helps understanding. The sentences added on page 2 and at the beginning of the index now indicates that it replaces a glossary (see the suggestions below).

-- The Fifth Option, the "Service Bureau" Alternative (Villanueva, 18 March, point 4): This is mentioned in passing (pg.5 point 4, and pg.35 last paragraph). However, I agree that it SHOULD RECEIVE MORE EMPHASIS. It seems best to consider it as an a variation of the first computerization option in Section IV. I have introduced appropriate sentences wherever relevant; the suggested modifications are given below.

However, the specific problems of conflicting computer systems in the service bureau and the parent agency are NOT considered, since they involve issues not relevant to most Population Information Centers and too complex and too far outside my experience to be included without extending both my knowledge and the Guide.

-- NETWORKING (Villanueva 18 March, point 5): It is difficult enough for a small, low budget, Population Information Center to automate, and it would only complicate matters more to discuss technical problems of database conversion, etc., which are not relevant to most Population Information Centers now. Therefore, this topic is NOT INCLUDED.

In any case, the emphasis placed in the Guide on compatibility with the UNISIST Reference Manual and the use of descriptors from thesauri recommended by POPIN will insure that the few Population Information Centers that become involved in tape exchanges can make the conversions relatively easily.

-- Photographs (Villanueva 18 March, point 7): They would be nice but I doubt the cost of photographs would warrant their inclusion.

II. Specific comments and suggestions, in order of page number
(Suggested changes in the text are indented.)

-- Page v (Contents). From the top of the page:

- "8. Data entry, correction and modification
- 9. Database security
- C. Decisions concerning output
- 1. Searches
- 2. Production of journals and listings
- D. Decisions concerning hardware and software
- 1. Computer/software combination available
- 2. The microcomputer/software solution
- 4 Peripheral devices

VII. UTILIZATION OF A MANUAL SYSTEM TO FACILITATE AUTOMATION ."

(rest without change)

-- Page 2, Note 3 at bottom of page. Change to read:

"3/ A term is underlined in the text whenever it is defined explicitly or used in a context that helps explain its meaning. Rather than provide a Glossary, an Index is given at the end of the Guide to facilitate finding the definitions of the underlined terms."

-- Page 4, para 1, line 2. Should read: "has various disadvantages", not "have".

-- Page 4, Automation option 1: Change to read:

"1. Automate the storage and retrieval of bibliographic information as soon as feasible, for which there are two basic alternatives:

A. Automate "in-house";

B. Use a "documentation service bureau" to which bibliographic information can be sent for storage and from which outputs can be obtained."

-- Page 4, Automation option 4. Change to read:

"4. Obtain searches from computerized databases elsewhere."

-- Page 4, Recommendations at bottom of page: Interchange points 1 and 2, since expert advice on system design is the most general point and should come before specific points.

-- Page 5, point 6 at top of page. To follow the grammatical structure of the other points, change to:

"6. The data entered into the computer should be organized for efficient searching..."

-- Page 5, point 3 at bottom of page. Starting on line 4, change the phrase to read:

"...Research Centre (IDRC), Ottawa; these software systems, suitable for manipulating large databases, are free and the corresponding agencies will provide..."

-- Page 5, point 3 at bottom of page. After "ESCAP is providing technical assistance to user centers in the region.", add the content of the phrase suggested by Kolbe:

"If the Population Information Center has access to a computer different from these in the parent agency, it may be able to locate and obtain suitable software from among other users of that computer model."

-- Page 5, point 4 at bottom of page. Change to read:

"4. If the above computer/software combinations are too costly, the Population Information Center should postpone automation until a low-cost microcomputer with appropriate software can be obtained or until an adequate local agency can provide reliable computerized input and output services for the Center at reasonable cost."

-- Page 6, point 5. The first sentence is so similar to that of Point 4 that it has to be read twice to see the differences. It could be changed to read:

"5. For a Population Information Center, with a small database, a limited budget and no previous automation experience, an in-house microcomputer system is likely to be the most practical approach to computerization in the long-run. The software system selected should be user-friendly and be ready to be utilized for the storage and retrieval of the bibliographic information without assistance from a computer programmer."

-- Page 5, point 4. Eliminate the sentences in the margin "However, if ... to proceed with automation.", since the idea is now incorporated in point 3.

-- Page 6, points 6 to 8. These points are too specific to be included as recommendations in a Executive Summary. Therefore, points 6 through 8 SHOULD BE ELIMINATED from this section, although their content is now included in the main text and summarized in point 5 on page 6.

In addition, the information that I have on the CONDOR system is that it is a multi-purpose database system and that one would have to program CONDOR to use it for a bibliographic application. Hence, unless there is a CONDOR program available for the storage and retrieval of bibliographic information that is reasonably compatible with the UNISIST Reference Manual, it is not likely to be very useful to a small Population Information Center without experience. Section VI.D.2 now mentions CONDOR and what seems to be a much easier system to use, CARDBOX.

-- Page 7, point 2, third line. Change to read: "As of January 1983, four such services ...", instead of "three".

-- Page 7, point 2 (c). Change to read:

"(c) CELADE/DOCPAL, with around 19,000 records containing abstracts in Spanish on Latin American and Caribbean documents and an annual growth rate of around 1,500 to 2,000 records, provides free searches and a document copy service; and (d) EBIS/POPFIL, the..."

--Page 8, para 1: Change to "five basic hardware..." instead of "four".

-- Page 8, point (a). Although many texts describe a computer as I indicated, it may be convenient to use the form suggested by Valantin (31 March, pg 2). Therefore, change point (a) to read:

"(a) Central processing unit. This component, often called the CPU, consists of registers, which retain a very limited amount of information actually being processed, immediate results and data on the status of the operations in progress, a control unit to interpret..."

-- Page 8. After point (a), add:

"(b) Main memory. This high speed storage, now on chips, holds such information as intermediate results and the step-by-step instructions for carrying out the operations on the data of interest."

-- Page 8. Starting with "Auxiliary storage" change the letters of the points to (c), (d) and (e).

-- Page 10, Figure 1. See attached Figure.

-- Page 11, last para. Since everyone I have ever met calls them "IBM cards", change first sentence to read:

"...may be done using 80 column cards (known as Hollerith or IBM cards), it is increasingly..."

-- Page 12, Para 2 (Wild 23 February). Add the following sentence to the end of the paragraph:

"While physical links by a cable from the terminal to the computer may be able to use cheaper modems, the use of telephone lines can be more flexible since with appropriate modems the terminal can be connected wherever there is a telephone."

-- Page 17, Figure 4. As was suggested for the Inverted File Search, the search statement and explanation for the Serial Search should be brought up above the diagram:

"Search statement
<Peru> AND <Colombia>

Read each record."

(DIAGRAM)

-- Page 18, last paragraph. To correspond better with the examples in Annex II, change the last three sentences to:

"...the effect of infant mortality on fertility in Colombia and Peru, and discuss..."

-- Page 19, points 5 and 6. I think I thought that the larger number of access points and greater retrieval facilities in a computerized system would help deal with fugitive documents that often have incomplete information on their author, etc. This is not a very important, so delete point 5 and change point 6 to 5.

-- Page 20, Automation option 1. To take into account Villanueva's (18 March) Fifth Option, change to:

"1. Automate the storage and retrieval of bibliographic information in the Population Information Center as soon as feasible, for which there are two basic alternatives:

A. Automate "in-house";

B. Use a documentation service bureau, e.g., another center with a computerized database, to which the Population Information Center's bibliographic information can be sent for storage and

from which outputs can be obtained."

-- Page 20, The last two paragraphs should be combined:

"4. Obtain searches from computerized databases elsewhere. This option, which allows any population information centre to obtain the benefits of computerized searches, is clearly not in conflict with the other options and should be considered whenever appropriate even in computerized Population Information Centers."

-- Page 22, para 1. The 1st paragraph with the inserted sentence should be rewritten as follows to avoid confusion:

"If the users are primarily policy makers, high-level administrators and similar persons who normally require consolidated information, a computerized system can be an important tool to aid population information analysis centres in preparing the consolidated documentation required when the amount of primary information is large and varied. It is unlikely, however, that the policy makers will utilize the database retrieval facilities directly."

-- Page 23, after second paragraph and before Section E. Following the suggestion of Pasquariella (15 March, para 3), insert the following paragraph:

"When developing a computerized documentation system that will serve external users, the Population Information Center should consider the simultaneous development of a document delivery system as well. Although document delivery is not part of the computerized system, per se, the greater facility with a computerized system to identify items of interest is likely to bring a greater demand for copies of the items not easily found in local libraries."

-- Page 26, para 1. Replace the 1st paragraph with the following two:

"It will be assumed here that the decision has been made to create a computerized database (first automation option stated in Section IV) and that the services derived from the database will be searches on request and the production of a journal (see UNESCO, 1980b for a guide on using a computerized database for the selective dissemination of information). If a reliable documentation service bureau is available and meets the needs of the Population Information Center at a cost it can afford, this alternative should be carefully considered since the Centre then will be relieved of dealing with many of the more technical aspects of automation outlined in this section and will be able to concentrate from the beginning on using the computerized facilities for the improvement and extension of its services.

"On the other hand, if the "in-house" automation alternative is to be undertaken, a number of key system design decisions described below must now be made on substantive, documentation and computer matters. It is

important to remember that these are highly interrelated and that many of the substantive and documentation decisions will depend ultimately on the software that can be obtained (at a reasonable price) for the computer equipment to which the Population Information Centre has access."

-- Page 26, para 2. As the actual and potential values of microcomputers for Population Information Centres have been mentioned up to this point only in the Executive Summary, the first sentence of this paragraph should be changed to read:

"Although automation may soon become much simpler through the use of microcomputers and associated bibliographic information software, since their use is rather at an experimental stage at present, population information centres..."

-- Page 26, paragraph in the margin after point 2. This very specific point on dataentry is not appropriate to be included here where very basic advice is being given. Therefore, eliminate it here (see suggestion for page 33).

-- Page 30, para 1. Underline EBIS/POPFIL since it is now in the Index.

-- Page 31, suggestion in the margin on the POPLINE worksheet. The purpose of the Guide should be to provide concrete examples within a general set of rules that Population Information Centres beginning to automate can use to get started fast and to facilitate compatibility among centers. Ideally, there should be a single recommended approach that each Center can adapt to local needs and conditions; the cafeteria approach only confuses. Unfortunately, there is no agreed upon POPIN worksheet, so examples from working systems are given.

If the POPLINE worksheet is reasonably compatible with the UNISIST Reference Manual (I do not have a recent version), and fits within the general approach given in the Guide, it could be included. Otherwise, it should be excluded since it will not be clear to the users of the Guide how it relates to the rest of the text.

-- Page 31 (Valantin, 31 March, last point on page 4): Add the following paragraph after the last paragraph on the page.

"Although it is advisable that a worksheet be used by a Population Information Center which has no previous experience with automation, it is not mandatory and, indeed, in the long-run it should not be necessary to carry out the keyboarding operation twice (i.e., first writing or typing information onto the worksheets and then transferring the same information from the worksheet into machine-readable form). In the near future, the use of microcomputers will allow electronic worksheets to be filled in with various checking and other operations carried out automatically by the machine."

-- Page 32, last para., 3rd line. Change to read:

"number. As a variation of the document service bureau alternative to automation (see Section IV), it may be possible in some cases for...."

-- Page 33, para 2. Add the following sentence to the end of the paragraph:

"As noted in the discussion of worksheets, microcomputers can be used effectively in many cases for data entry and editing functions; this is a reasonable approach even in large computerized storage and retrieval systems."

-- Page 33, last paragraph (Inverted file). While I agree this (and the section on Searches, page 34) should be moved from their original locations, I do not think it should be moved to page 40, mixing it with decisions on devices, etc. Rather, I propose that it and the paragraph on Searches be moved to page 36. See the suggestion for page 36 below.

-- Page 34, para 2 (Searches). As noted in the previous comment, this should be moved to page 36, rather than page 40. See the suggestion for page 36 below. However, since searches are such an important part of output, some mention of them must be made under the subsection on Output. I suggest the following:

"1. Searches

As searches requested by users on specific topics are likely to be a major form of output, careful thought must be given to the form of presentation of the information retrieved so that the user can interpret the citations and other data without assistance. If the number of records located is large, it may also be necessary to consider the most appropriate ordering (e.g., by author) of the printed output from the computer. Although these decisions are substantive, they depend upon the capability of the software to do what the documentalist requires. For additional comments that affect software decisions involving output, see Section D below."

-- Page 34, Scratched out title to the last paragraph. This should be included since it like "searches", should be emphasized as an important area of output. Hence, the section subtitle before the last paragraph should be:

"2. Production of journals and listings"

-- Page 35, last paragraph, 2nd line. To be consistent with the previous suggestion, remove the word "interactive", and rewrite so that the phrase reads: "...countries will want to conduct searches on demand and produce listings..."

--Page 36, para. 3. Change the 8th line on the page to read:

"...delays (see section D.2 below for microcomputer approaches that are likely to soon be available to Population Information Centres in developing countries)."

-- Page 36, para. 3. Insert before the section on computer/software combinations:

"From the discussion in section II above and the description of Boolean Algebra for searches in annex II.C, it should be evident that better service can be given if the software selected permits interactive searches, since the searcher can "learn" from an examination of the information being retrieved and make modifications immediately to come closer to the user's needs. However, batch searches should also be possible for making large standard lists, etc. The software should be user-friendly, in that it should be easy for a documentalist or a motivated user to learn and should not require assistance from programmers.

"It is very important that the software be able to make searches on an inverted file (see section II) and it is desirable that the selected fields of new records entered into the master file be included in the inverted file without having to reinvert the entire file. This may be fully automatic inversion where the selected fields are inverted as each new record is entered into the database or it may involve storing new records until there are sufficient to carry out a special process to add the new records to the inverted file."

-- Page 36, 2nd para from the bottom. Change the phrase starting on the second line of the paragraph to:

"...to non-commercial software, that is essential free, for the mini- and mainframe computers..."

-- Page 36, last para. Change the first line to read:

"There are two major non-commercial software systems available, that can be used for large databases: ISIS and..."

-- Pages 36 and 37, last para and Table 3. The paragraph and the Table refer to non-commercial software because it is essentially free, a sine-qua-non for most small Population Information Centres which cannot afford the thousands of dollars for rental or purchase of commercial database systems. I have no information on the cost of the IRS-4 system used by ESCAP on their NEC, although since it is apparently a commercial product, it may be too expensive for most Population Information Centres. If it fits the criteria given and is low-cost it should be mentioned and included in Table 3.

--Page 37, Table 3, fourth IBM computer system. Change to read:

IBM Mainframe (a)	OS for batch using CDS/ISIS; CMS for on-line operation	ISIS	CEPAL, Santiago
-------------------	--	------	-----------------

-- Page 37, Table 3, last computer system (UNIVAC). Eliminate and eliminate note (c).

--Page 37, Table 3. The "Corinthe system" for thesaurus manipulation should NOT be included since it is not an bibliographic storage and retrieval system for carrying out the operations described throughout the Guide.

Information on the Corinthe system should be given in the POPIN Bulletin and the system should be available on tape along with a file of the entire POPIN Multilingual Thesaurus.

-- Page 38. Insert after the first paragraph:

"If the Population Information Centre has access to a mainframe or mini-computer other than IBM or HP-3000, it could obtain information from among users of the particular computer model on existing software for information storage and retrieval. If adequate software is located, it may be possible to proceed with automation.

An increasing number of NEC mainframe computers, which can employ the commercial IRS-4 information storage and retrieval software, are in use in Asia. ESCAP, which utilizes this computer/software combination for its EBIS system, provides technical assistance to users in its region."

-- Page 38, sub-section 2. Since there are many changes to be introduced into this section it is easiest to eliminate the entire previous text of this sub-section and replace it with the following:

"2. The microcomputer/software solution

"The costs and various problems associated with the computer/software combinations described above will make them unsuitable for in-house automation in most Population Information Centers. Fortunately, the rapidly increasing capabilities of microcomputers and memory devices and decreasing cost of the equipment, coupled with the development of low-cost software, will make automation feasible for many population information centres with relatively small databases (perhaps up to 2000 bibliographic records).

One of the major areas of software development at present for microcomputers is database management systems; a DBMS is a generalized set of computer programs to control the creation, maintenance, and utilization of computer-based files. Ferris (1983) provides a list of over 25 DBMS available for microcomputers, with information on the

operating systems and microcomputers on which they work, addresses of suppliers and prices.

While most of the DBMS listed by Ferris are primarily oriented toward business database applications like personnel information and inventory management, some, like the CONDOR Series 20 software (priced from US\$295 to \$650), can be employed for storing and retrieving bibliographic information. However, setting them up for a bibliographic storage and retrieval system may require the services of experienced computer programmers.

A software system that comes ready to be used for bibliographic information is CARDBOX, which costs US\$245 and can be used on a wide range of microcomputers. It has very easily understood and short manuals and simple commands that permit users to set up a database, add new records and make corrections, conduct interactive searches using an inverted file, print results and carry out many other operations. Although CARDBOX could be extremely useful for Population Information Centres with one or more bibliographic and other databases, it cannot be used with lengthy abstracts or with all the UNISIST Reference Manual essential fields, because the maximum number of characters per record are 1404 and the maximum number of fields is 26.

UNESCO is supporting the development of a system known as IMD IV+V 4/, since the system will be set up to be used with the complete UNISIST RM. To allow it to be portable among a wide variety of micro, mini and some mainframe computers, it is being written in the widely-used high-level programming language, UCSD PASCAL 5/ by the Institut fuer Maschinelle Dokumentation (IMD), a non-profit research and development corporation in Graz, Austria. Most of the software development work is done on PDP minicomputers (PDP 11/2 and PDP/23) and the first implementation of the software is on a Xerox 810 microcomputer (Park and Stearns, 1982).

The cost of a microcomputer and associated equipment will vary widely from country to country, but is likely to be in the range of US\$10,000 to \$15,000. For example, an Apple II Plus with a floppy disk drive, a hard disk drive, a video screen and backup equipment costs less than around US\$11,000. Similarly, a NCR Decisionmate V with a 10 megabyte hard-disk drive, a 320 kilobyte floppy disk drive and a printer costs between US\$9,000 and US\$12,000 depending on the importation costs. In many cases, the cost will be low enough to permit purchasing a second system that will both facilitate the work and act as a back-up computer should one fail."

-- Page 39, last sentence of point (a). I do not see why a typewriter terminal will be very useful for interactive searching except that it gives hard copy. Since a typewriter terminal is very slow in comparison with a video terminal, it is quite painful to use when carrying out interactive searches. Change the last line to read:

"Typewriter terminals for interactive work should be avoided; however, some form of printer, which can be a typewriter terminal, is required for obtaining the final results of interactive searches."

-- Page 40. Eliminate the insertions after point (e), since the information on these is now included on page 36 (see above).

--Page 41. Comments in the margin: The purpose of the system described is to make it possible to carry out searches to identify documents using simple Boolean algebra expressions of descriptors, geographical areas, etc; i.e., unlike an ordinary card catalog, the system permits computer-like searches. Numbers have to be used since they can be arranged in columns by their last digit to facilitate rapid visual comparisons on the UNITERM cards. After determining the number of each document which matches to the specific combination of descriptors, etc., one must go to the Master cards ordered by number to determine if each document really is of interest.

If there is much use for a separate author file, the Master file cards could be duplicated to order them by author. While such a file would eliminate the "inverted author" file on the UNITERM cards and might be a very good idea, the author master file cards would not assist the user in a subject oriented search which should be the main concern in most Population Information Centres handling external queries. I would not recommend the UNITERM or similar-purpose system to a library primarily serving internal users requiring traditional library services.

Hence, unless there is some point that I have missed, I would suggest changing only the last sentence of para 4 on page 41:

"The population information centre can create a manual "inverted file" for each field of interest, such as descriptors and geographical areas treated. For the simple location of documents by particular authors it may be useful to duplicate the master file cards and order the second set by author."

--Page 66 (Bibliography). Insert after the Dierickx citation:

"Ferris, David. 1983. Micro DBMS: The little engine that can. Computer Decisions. Vol 15, No.5 (May 1983): 152-162."

-- Page 68ff (Index). The following terms should be added (page numbers refer to the handwritten numbers in the Pop Div text):

Database management system 38
DBMS 38
DOCPAL (see CELADE/DOCPAL)
Document service bureau 4, 20
EBIS/POPFIL 31, 38
Hollerith cards 11, 15
POPFIL (see EBIS/POPFIL)
Registers 8