

*Brian W. Carss*  
*Director of ERIC Clearinghouse on Early Childhood Education*  
*University of Illinois, Urbana*

## THE DEVELOPMENT AND OPERATION OF AN EDUCATION RESOURCES INFORMATION CLEARINGHOUSE—(ERIC)

In 1959, the Office of Education became particularly interested in developing an information retrieval and dissemination system for research that was carried on under Federal sponsorship. A pilot study was authorized to demonstrate the feasibility of such a system within the field of education. It was not until the years 1965-66, however, that the first of the twelve Education Resources Information Clearinghouses (ERIC) was established, and not until later, in March 1967, was the ERIC Clearinghouse on Early Childhood Education established as a part of a new program to be sponsored, again by the Office of Education.

The ERIC Clearinghouse, in this latter instance, was founded as part of a programmatic effort to stimulate research and development activities in early childhood education. This programmatic effort is called the National Laboratory on Early Childhood Education. The Laboratory is a consortium of six universities: the University of Arizona, the University of Chicago, Cornell University, Syracuse University, and George Peabody College at Nashville, Tennessee. The efforts of these five research and development centers are coordinated by a National Coordination Center at the University of Illinois. The ERIC Clearinghouse is an integral part of this National Coordination Center.

Figure 1A shows the ERIC Clearinghouse at the hub of a wheel which includes preschool research and development activities. The wheel is divided into three portions, research and development, implementation and dissemination, and information processing. Within the research and development sector are all the activities of the National Laboratory. Figure 1B in the diagram

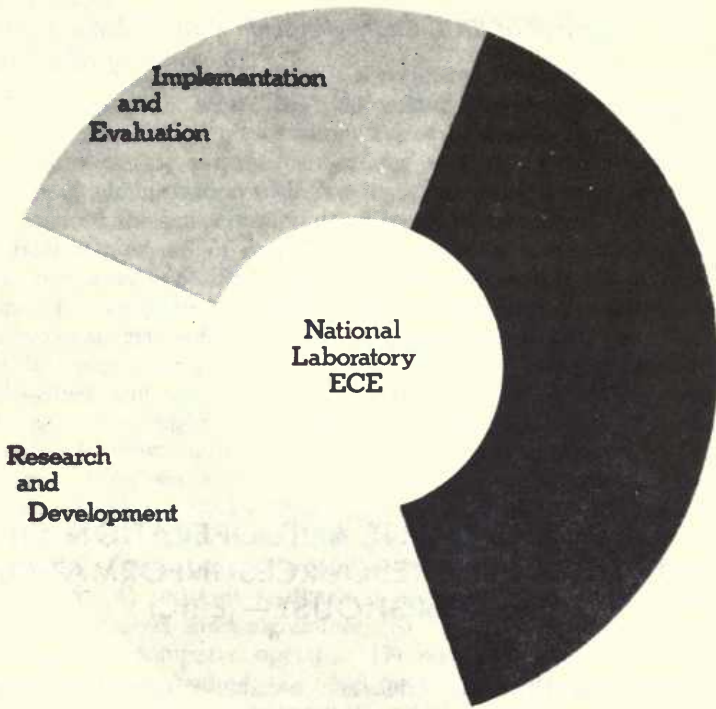


Figure 1A. Activities Related to the ERIC Clearinghouse on Early Childhood Education

shows how information flows among the research and development centers, which have a responsibility to transmit their research findings to the Coordination Center so that these findings can be disseminated by the ERIC Clearinghouse to the education community at large. However, the research and development centers themselves do not just transmit information one way. They can also receive information from the ERIC Clearinghouse.

In the implementation and dissemination sector are three major groups. These are the Head Start groups, the Follow Through groups, and the Parent-Child Centers. The Clearinghouse has a prime responsibility to the Office of Economic Opportunity (OEO) because OEO funds the ERIC Clearinghouse operation by granting the money to the Office of Education to administer. The ERIC Clearinghouse, therefore, apart from operating a basic clearinghouse is obligated to perform special services for Head Start. In a similar way, the clearinghouse performs special services for the Follow Through programs and the Parent-Child Centers. The kinds of services that ERIC is being called upon by Head Start personnel to perform include

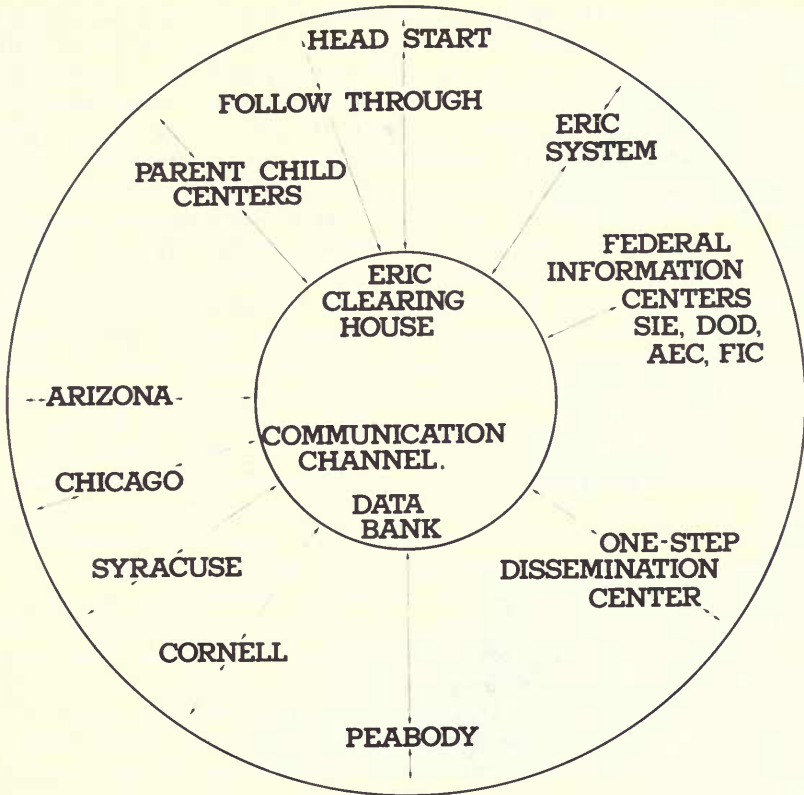


Figure 1B. ERIC Clearinghouse Relationships with Other Information Sources

generating an insert for their newsletter every other month and commissioning the writing of pamphlets or non-technical bulletins for distribution to people in Head Start programs.

The third sector, that of information processing, includes all the basic operations of an ERIC Clearinghouse. The basic Clearinghouse operations include acquisition and processing of documents, and performance of some information analysis. The Clearinghouse has developed working relationships with a number of other information sources. As shown in Figure 1B, these include the Office of Education, Scientific Information Exchange, the Federal Science and Technical Information Clearinghouse, and the major organizations that cater to the field of early childhood education. Finally, a source of information that is somewhat harder to tap includes all the people who are working in the field of education of young children. These are people such as kindergarten teachers, school boards, superintendents, and community action program teachers.

One aspect that has not been discussed so far is the relationship between the ERIC Clearinghouse and the Coordination Center of the National Laboratory. These two operations enjoy a symbiotic relationship and the Coordination Center, of course, can use the ERIC Clearinghouse as a communication channel to expedite the transfer of information between its five Research and Development Centers, as shown in Figure 1B. The communication channel handles all items that are in preliminary form such as preprints of papers, untried or partially tried curricula, etc. Such items are distinctively marked so that, for the moment, they will not be widely disseminated. People who receive copies of these materials are asked for critical comments which can then be channeled back to the authors. This critical feedback is beneficial, and an essential part of all developmental work. The Clearinghouse, in return, relies upon the National Laboratory for substantive support. In the National Laboratory, particularly in the Research and Development Centers, the people are, hopefully, at the forefront of knowledge in early childhood education and therefore are particularly aware of the needs for information in the area.

Figure 2 shows the approximate state of the ERIC Clearinghouse on Early Childhood Education after one year of operation. As an aside, March is not the best time of the academic year to begin funding of any operation. However, we have attempted to capitalize on the quiet of the summer months to initiate a public relations program and to advertise the existence of the Clearinghouse on Early Childhood Education. Fortunately, the high proportion of teachers and school personnel on campus in the summer provided us with a select audience. The Clearinghouse staff stood outside classrooms as people entered and left, and distributed descriptive literature on the scope of work, and charge of the Clearinghouse. It is very difficult to quantify how successful something like this is. To gain an estimate, we walked down the corridor after everyone had gone and noted how many of the sheets were lying on the floor or on window sills, or had been thrown into trash cans.

It was not until September of 1967 that the operation of the Clearinghouse got fully underway. In November 1967, we submitted our first set of documents for inclusion and announcement in the monthly bulletin *Research In Education*. Each month since then we have submitted approximately twenty-five fully processed documents for inclusion in this bulletin.

## THE ERIC SYSTEM

### Document Processing

For those who are unfamiliar with ERIC, Figure 3 presents a brief outline of what is involved in processing a document for inclusion in the system. As documents arrive in the Clearinghouse, they are given a document routing card and a unique accession number. The accession number consists of the Clearinghouse code letters and six digits, for example, PS 000241. This number is stamped in the lower left hand corner of the first page of the document and at the same time it is stamped on the document routing card. The document routing card includes descriptive information about the document (such as the document title, when it was received, and the source of the

STATE  
OF THE  
CLEARINGHOUSE  
AFTER ONE YEAR  
OF OPERATION

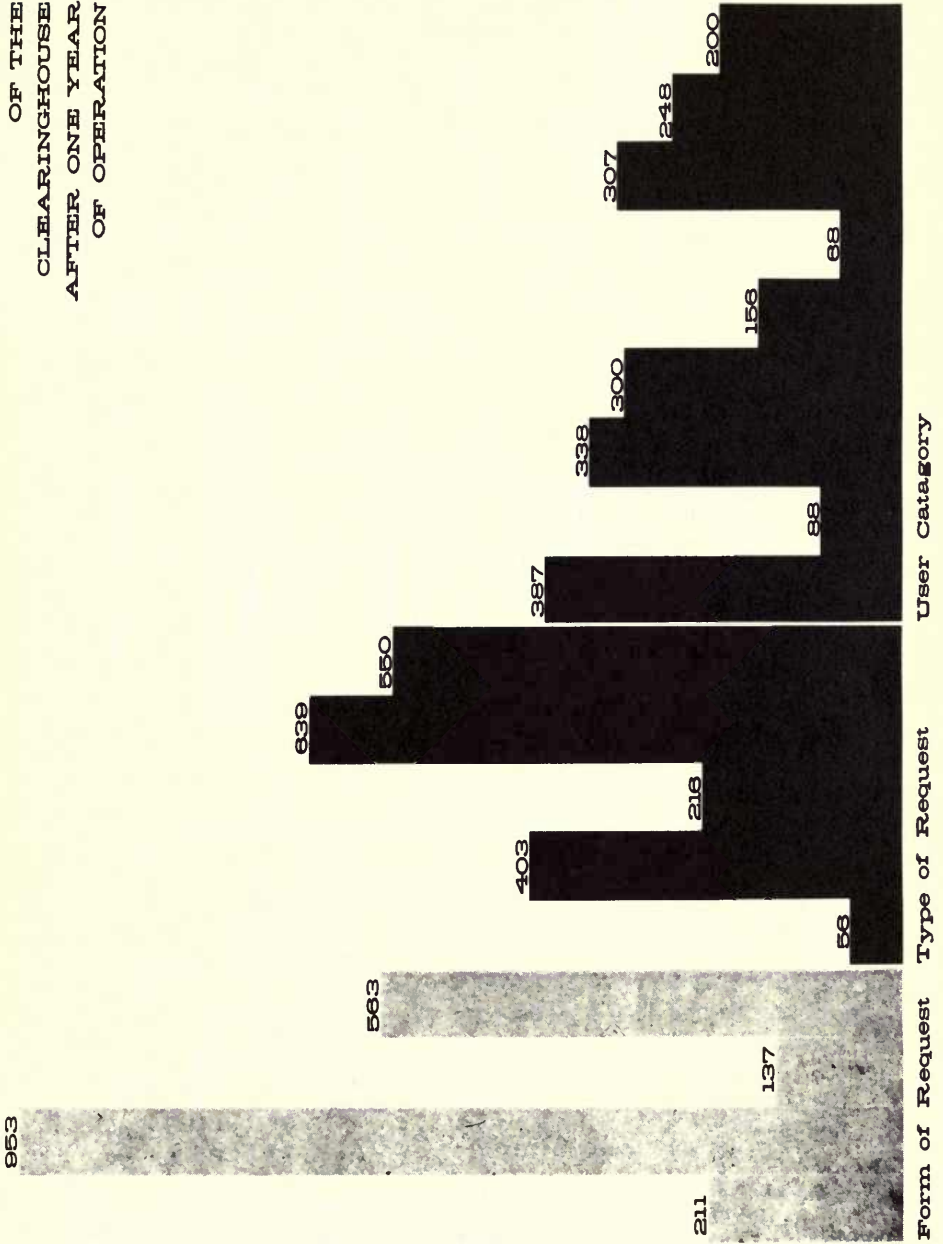
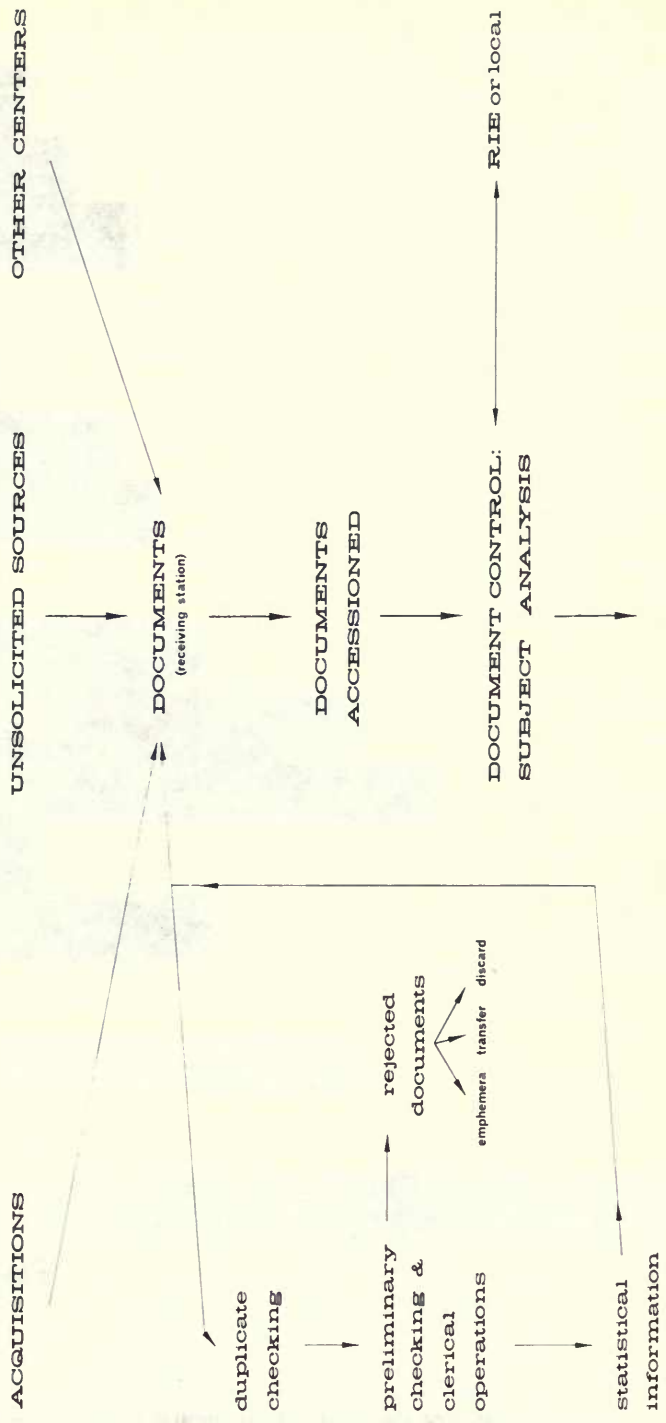


Figure 2. State of the ERIC Clearinghouse on Early Childhood Development after One Year of Operation



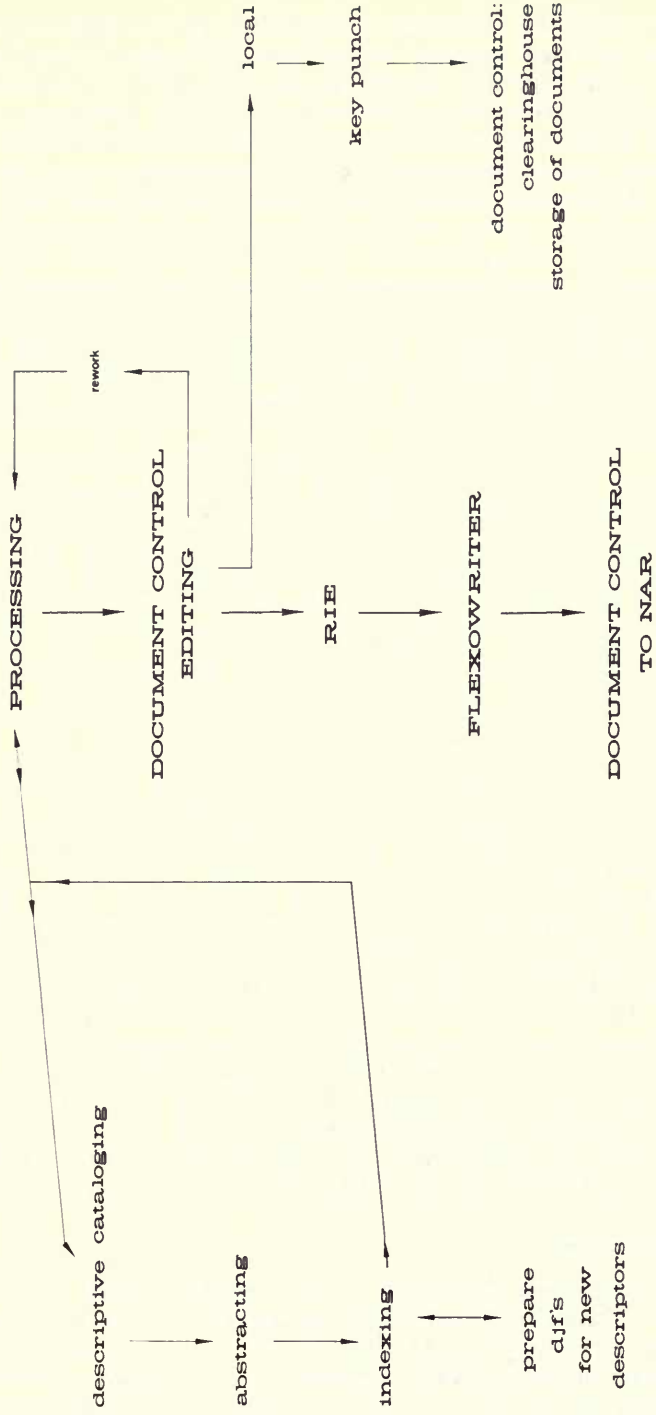


Figure 3. Processing a Document for Inclusion in the ERIC System

document), and as the document proceeds through the processing, various other dates are filled in, such as the date it is indexed, the date it is abstracted, and the file date for submission to *Research In Education*. The documents are screened for quality and relevancy by evaluators who are people with background experience in early childhood education. The evaluating procedure assigns a document to one of three major groups. Group One documents are recommended for inclusion into the National ERIC system files. These are high quality research documents. Group Two documents are held locally in the Clearinghouse files. The latter documents may already have been processed by other clearinghouses and entered into the system, or they may be good documents but not of the highest quality. It is very necessary for a clearinghouse to hold all such applicable documents for the sake of completeness. The third group of documents are rejected either because they are not relevant to the domain of early childhood or are of such poor quality that they can not be used.

After a document is evaluated, the abstractors in the Clearinghouse do the descriptive cataloging, indexing, and abstracting. Descriptive cataloging involves recording the author, title, institution, source, sponsor, grant number, and prices for microfiche and hard copy on the standard ERIC resume form. Indexing involves the extraction of terms that best describe the document. By using a set of these terms, the computer files can be searched automatically to retrieve the appropriate document and ones similar to it. The document abstracts are approximately 200 words long and generally indicate the problem investigated, how it was studied, and the general conclusions. The term abstract is perhaps not a good one to use when referring to an ERIC resume. A much better term, and certainly much less confusing, would be announcement. If, after being screened for quality, the document is recommended for inclusion in the national system and all the Clearinghouse processing has been completed then the document and its resume are sent to the ERIC System's computer processing facilities at North American Rockwell in California. North American Rockwell currently has a contract with Central ERIC to perform the document computer processing. Here they are added to the ERIC master files. A print-out of these newly added announcements is photographed for inclusion in the U.S. Office of Education's monthly bulletin, *Research in Education*. The document, if a copyright release has been secured from the author or the publisher, is given an ERIC document acquisition number. This is very similar in form to the Clearinghouse number. It consists of two letters, ED in this case, and six digits such as ED 001432. The document is then sent to the ERIC document reproduction service where it is photographed and put onto microfiche. These microfiche may then be obtained by anyone for a nominal fee. Hard copies are also available; however, they are more expensive.

### Information Analysis

The Clearinghouse functions that have been described up to now are all part of the operation of a basic ERIC Clearinghouse. Another very important function that must be performed is that of information analysis. Information analysis means the interpretation, synthesis, and repackaging of research results and other information for a variety of audiences. This analysis make take



the form of state-of-the-art reports which discuss the key trend of research in specific areas of early childhood education, or it may take the form of current awareness articles or bibliographies. As already mentioned, this Clearinghouse is also obligated to produce a number of non-technical pamphlets describing early childhood research related to issues confronting Head Start. These pamphlets are an additional part of our information analysis activities. Four monographs and a similar number of current awareness articles have been commissioned for this year.

### Internal Retrieval System

The informational demands that are already being made on the Clearinghouse indicate that the decision to operate a computer-based retrieval system was correct. For an interim period, the basic information retrieval system developed at Michigan State University is being made operational on the University of Illinois' IBM System 360 50/75. The package of programs is very flexible and allows a user to make optimum use of the machine's capabilities and to use these capabilities in supplementing his own. For example, in document processing no way as yet has been developed for automatically indexing and abstracting a document, and so it is more reasonable for these operations to be performed by a human being. On the other hand, the human being cannot as a rule compete with the speed of a machine in carrying out a search of a master file, especially when the file is large, it is therefore more reasonable in such cases for the searching and retrieval of information to be carried out by a machine. However, there is also a middle ground where the interaction of a man and a machine produces a much better product than if either a man or a machine were allowed to do the work alone.

The basic information retrieval system consists of a number of computer programs which operate under an executive program. This enables the user to organize the system into a number of independent clusters of dependent operations. For example, flexibility is built into the system by:

- 1) completely automating only the clerical operations common to most information retrieval systems, i.e., file maintenance, up-dating, and searching;
- 2) allowing the user to control the operations which are highly dependent upon his own particular retrieval problem, i.e., indexing and classifying documents; and
- 3) providing the user with a choice among different methods of indexing and searching the information.

The model that has been used to construct the program package can be split into three types of information elements: notably abstracts, descriptions, and questions. Three fundamental operations are associated with these elements: information storage, information indexing, and information retrieval. These basic operations may be clarified by analogy with traditional library operations. The information storage corresponds to the maintenance of the collection of books and periodicals in the library; the information indexing corresponds to the maintenance of an index to the contents of the library, such as author, title, and subject catalogs; and the information retrieval operations are analogous to the process of searching author, title, or subject

catalogs for the call numbers of relevant documents and then retrieving the documents themselves by means of their call numbers.

The retrieval system includes a separate set of computer programs for each of the three fundamental operations of information storage, indexing, and retrieval. And each component program is designed to help the user with a single fundamental operation. To construct a particular type of information retrieval system the user simply selects the proper component programs on the correct sequence. The main advantage of this so-called modular design, of course, is flexibility. Since the fundamental operations are mutually independent, the component programs may be used or modified independently. Hence, different components may be combined in various ways to produce a variety of information retrieval systems. Similarly the entire system may be selectively modified by replacing existing components—or adding new ones, for that matter—to the fundamental operations without any modifications to the components already in the system.

The following table summarizes the operations performed by the basic information retrieval system.

**TABLE A**  
**Operations Performed by the Basic  
Information Retrieval System**

Operation	Performed by the User	Aided by the Information Retrieval System	Performed by the System itself
Information storage/abstracting	Yes	No	No
Information file maintenance	Yes	Yes	Yes
Information indexing/analysis	Yes	Yes	Yes
Description file maintenance	Yes	Yes	Yes
Information retrieval/questions	Yes	No	No
Searching	Yes	Yes	Yes
Retrieving	Yes	Yes	Yes

### Applications of the System

For the purpose of the present discussion, we shall describe only two applications: the development of a printed author-title catalog, and an automated reference service for the test library.

The printed author-title catalog is essentially a book index describing all documents included in the library. The catalog includes abstracts of all documents with a traditional author index and subject index produced by selecting key words from each abstract. The steps involved in generating such an index are summarized in Figure 4. As shown in the figure, the first step is

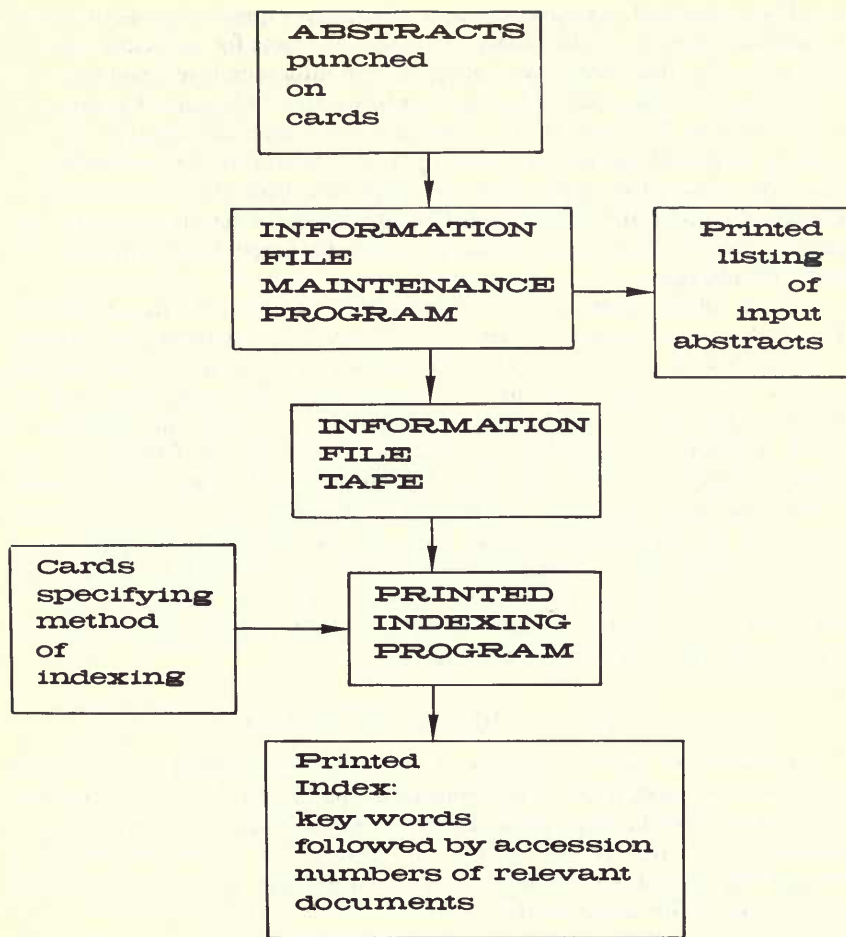


Figure 4. Generating an Index from Abstracts

to store all the abstracts on an information file tape using the information file maintenance program. This program reads the abstracts from cards, assigns unique accession numbers, and stores these accession numbers and abstracts on tape. In this application the program is also used to generate a printed report containing the set of abstracts ordered by accession number. The second step is to use the printed indexing program to generate and print the key word index. This program reads control cards which specify the method of indexing and then uses the abstracts and accession numbers on the information file tape to generate the index containing the alphabetized list of key words. Each key word is followed by the accession number of the relevant text abstracts. The printed set of abstracts and the key word index are combined to form the book index.

The automated reference service is essentially a question-answering facility which allows users of the library to request abstracts for particular types of documents. The first step in generating such an information retrieval system is to store all the abstracts on the information file tape by means of the information file maintenance program. Next, each abstract must be indexed or described by means of a set of key descriptive terms punched on cards. The descriptive cards may be generated manually by the user or automatically from the information file tape by the descriptive analysis program. In either case, these cards are used to prepare an index on the descriptive file tape.

Users submit questions to the system in the form of key terms describing the type of documents they are interested in. These terms are punched in cards and read by the descriptive file search program, which compares each question statement with all the test descriptions previously stored on the descriptive file tape. Control options are provided to permit the selection of various methods of narrowing or broadening the search. Unless the user specifies otherwise, a similarity index is calculated on the basis of the number of matching terms between questions and descriptions, and the accession numbers of the specified number of abstracts are selected by the program on the basis of this index of similarity. The question and accession numbers for relevant abstracts are stored on the question file tape. They can then be printed out as a report for each question, which includes all the relevant abstracts ordered by the index of relevance.

## EVALUATION OF THE SYSTEM

Towards the end of last year, the Clearinghouse submitted to a review of its operating procedures. This is something that all the ERIC Clearinghouses were asked to do in response to Federal concern with cost effectiveness. In preparation for this review session, an analysis of the Early Childhood Clearinghouse operations was performed. The following outline of this analysis shows its five major components:

### Input

*Acquisitions.* The Clearinghouse acquisitions include published materials from journals and non-published ephemeral materials such as papers presented at meetings, letters between investigators discussing issues, speeches, etc. A category not covered previously, but a most important part of the Clearinghouse operation as a consequence of its close ties to the National Laboratory, is that of non-print materials such as films, videotapes, slides, magnetic tapes of language samples, and curriculum innovations.

*Analysis.* The analysis operation includes all the normal functions of the Clearinghouse which were described before, the functions of document processing, abstracting, indexing, and descriptive cataloging.

*Thesaurus Maintenance.* This is a responsibility of all the Clearinghouses. This category has been isolated because thesaurus maintenance and the filing of descriptive justified forms is a very time consuming and expensive procedure.

*Recording Results of Analysis.* This is a practice prescribed by central ERIC. In its present form, this tabulation process is almost exclusively for statistical information such as the number of documents acquired, processed and their disposition, the number of questions answered and for what specific user groups.

*Storage of Source Materials.* This is done for both print and non-print materials. It is feasible to consider storing print material in microfiche form, but the storage and duplication of bulky documents is unfeasible because of the limited space and money. This is not really a problem at the moment when the total in-house holdings are approximately 1,500 documents.

## Output

*Question Answering.* The second function is that of output, which includes all the answers to questions received in the Clearinghouse. The questions received can be grouped into two major groups. First are the current awareness questions originating either at Central ERIC or by an individual, and second are the retrospective questions which raise handling problems such as how far back in time the search should be made. For the present, a five year limit has been established. Questions can be further grouped together according to the nature of the institution, limitations by status of the questions, limitations by fields of interest of the user, and limitations that may be applied to questioners from a profit-making organization.

*Question Analysis.* The Clearinghouse has the prime responsibility for question analysis but as this is a very expensive operation, we are attempting to automate question analysis as completely and as rapidly as possible. Central ERIC does presently have a very limited capability of carrying out automated searches. Until the Clearinghouse is more highly developed it is necessary to shift much of the responsibility of the analysis to the user. Unless the user accepts much of the responsibility, there is a built-in basis for failure. Because of limited time and funds and the inadequacies of any thesaurus, the staff cannot presently satisfactorily handle all the demands.

*Search.* All the Clearinghouses are required to maintain an internal search capability. It may be necessary to assign an expert in information storage and retrieval to aid heavy users of the Clearinghouse. Such an expert would assist these heavy users with both question analysis and search.

*Delivery of Search Results.* Through time a number of canned bibliographies have been compiled which are helping to reduce the operational cost; in fact, at the moment, approximately twenty-two such bibliographies are available. Other items to be issued will include abstracts, additional bibliographies, and complete documents either in microfiche or hard copy form.

## Evolutionary System Development

To enable the Clearinghouse to perform more effectively in the future we are attempting to explore how to motivate the Clearinghouse user to reply

when he has received a service from the Clearinghouse so that some evaluation of services can be carried out. Two methods are being considered: (1) to limit future Clearinghouse service to users who give evaluative feedback or (2) make a charge for service unless the user replies.

We are attempting to develop a profile of Clearinghouse users, the types of questions asked, and uses made of the supplied information. Other criteria that may be applied to evaluate the performance of the Clearinghouse include:

(a) the relevance of sources supplied in answer to the user's questions;  
 (b) the quality of the abstracts (Did the abstracts enable the user to decide whether the full document was relevant?); and

(c) the quality of the documents. Before carrying out a rigorous examination of the abstracts it is necessary to look at the quality of the documents. All documents received in the Clearinghouse are presently screened for quality. A brief description of this procedure has been outlined above.

Since the present procedure leaves much to be desired, a study of document screening procedures will be carried out. Four or five reviewers from the major early childhood organizations will be invited to meet and they will attempt to formulate a set of guidelines that can be used in the Clearinghouse. The next step will be for the Clearinghouse to give the reviewers a test set of 100 documents and ask them individually to classify the documents according to the criteria they have developed. Disagreement among these individuals can be used to justify a non-rejection policy at the Clearinghouse level. Only those documents that satisfy all the criteria would be processed for entry into the national files. To determine the relevance of the substantive material, users of Clearinghouse services must be queried on these matters: the quality of the substantive material; the proportion or quantity of the supplied material that is relevant; the amount of relevant material that is missed; whether the full document was ordered and when the full document was read, and whether the document was relevant to the question being asked.

### Cost Analysis

The Clearinghouse operation can be broken down into seven sub-categories for purposes of cost analysis: (1) document acquisition, (2) processing of documents, (3) question answering and searches, (4) advertisement, (5) internal review and output evaluation, (6) output newsletters and other products, and (7) intra-national laboratory communication.

To date it has not been possible to put accurate cost figures on the document acquisition program; however, this should become considerably easier when the acquisition procedure has become better established. Processing of documents costs between \$12-15 per document. Handling of question answering services costs approximately \$2 per question, although, without qualification, this is not especially meaningful since it does not take into account the kind of search required. Current awareness questions have a low search cost but a high analysis cost, and vice versa for retrospective questions. Advertisement cost are not yet sufficiently well defined to give a good estimate; however, they are a relatively small item in the present budget.

Internal review of output evaluation costs approximately \$2 per question. Clearinghouse output costs vary considerably depending upon the

item. The bi-monthly newsletter has been averaging 10.5 cents per copy to produce, although no charge is made to the recipient.

A liaison has been developed with the University of Illinois Press whereby the Clearinghouse will supply the press with cameras-ready copy of its products in the future. The Press has agreed to print, advertise and market these products through its normal channels. This arrangement will allow the Clearinghouse to increase its information analysis activities by not having to divert its resources to printing, though the products will still be readily and economically available.

### THE FUTURE OF THE ERIC SYSTEM

Five problems have been identified as needing concentrated study in the future. The first is to determine the validity and reliability of the abstracting in the Clearinghouse. In such a study we assume that it will be necessary to have each document abstracted according to guidelines as laid down by Central ERIC, so that any abstractor would be able to produce an abstract that essentially describes the document in a similar way. It is not expected, however, that identical abstracts will result. IBM has carried out studies on the reliability of abstracting, for example, in which a number of their Ph.D's working in physical chemistry abstracted a number of documents in this field. The company then also hired a number of intelligent high school seniors who were primed as to the kinds of things that ought to go into an abstract and then given an identical set of documents to abstract. When the two groups had finished, their abstracts were compared and no significant differences were found. Such findings have led us to project as our second study the examination of alternatives to abstracting. In one such study, carried out at the Knowledge Availability Systems Center, University of Pittsburgh, document extracts were used, whereby the first and last paragraphs in a document were taken in lieu of an abstract. (If the last paragraph consisted of acknowledgements then the next to last paragraph was used.) This extracted information along with the normal document citation was then used to test the relevancy of retrieval for some 1,100 documents, and it was found that the extract method gave more relevant documents in retrieval than did abstracts of documents.

As study number three we are proposing an extensive analysis of users which will consider the users' status, qualifications, age, rank, and interests. Study number four will be the document evaluation study mentioned earlier. Study five, which is already under way, is an examination of the document acquisitions program. With regard to the latter, we may note especially in conclusion that central ERIC policy presently limits foreign document acquisition. However, in the field of early childhood education, other countries (including Switzerland, Israel, Sweden, United Kingdom, and Australia, as well as others such as Peru and Japan) are doing a considerable amount of research in this area which can not be neglected because it has already had a considerable impact on U.S. research.