OPERATIONAL COMPUTER-BASED SYSTEMS AT THE CHEMICAL ABSTRACTS SERVICE

Seldon W. Terrant, Jr., and J. L. Wood

The American Chemical Society (ACS) is a non-profit organization chartered by Congress as a scientific and educational organization. The Chemical Abstracts Service (CAS) is a division of the American Chemical Society.

Like any society the ACS is concerned with helping its members to get their papers published and to help chemists and chemical engineers to gain recognition in the scientific and general public communities. Equally important, however, the ACS wants to have available information put to use. To this end, it is concerned that all forms of research, development, and applied technology be fully recorded, indexed, and retrievable for the full scientific community. Chemical information is not used only by chemists and chemical engineers. It is used by many other kinds of scientists. In line then with the major purpose of getting available information put to use, it is clear that the ACS operates the Chemical Abstracts Service not only for the benefit of members of the Society, but for a broader reason. We want chemical and chemical engineering information put to use in the full scientific community so that the general public will reap the benefits which such use will provide.

When the full dimensions of the information problem for chemistry and chemical engineering are considered, it is easy to see why a community service is a requirement for all those who use scientific information. If information processing is carried out well, the information will be cheaper and more easily available. If it is not done well, such operations will not continue to exist. Considering not only chemistry and chemical engineering information, but also all of the other scientific and technical information available, it is not possible

---

Seldon W. Terrant, Jr. is Managing Editor, Special Publications and Services, Chemical Abstracts Service of the American Chemical Society.

J. L. Wood is Head Librarian, Chemical Abstracts Service.

for any one institution to meet all of its information needs from within its own resources. Community services such as Chemical Abstracts are a necessity if information is to be used to the best advantage.

When the full user community for scientific and technical information and the diversified needs of the community from institution to institution and from individual to individual are considered, it can be seen that the traditional means of publication, which assemble information into traditional printed form, are neither flexible enough nor responsive enough to meet today's needs adequately. For this reason, during the last five years the ACS/CAS started to shift its operations to a computer base. It is expected that all CAS publications and services will be composed through the computer by mid-1969.

In addition to providing acquisition, analysis, organization, and indexes for all chemical and chemical engineering information, CAS intends to provide systems through which the stored data can be adequately tapped for routine use by the scientific community. The shift to a computer base will provide the additional flexibility which is necessary to be sure that the information is fully used.

Several of the CAS computer applications involve or affect the operations of the CAS Library. This paper summarizes three of these applications. First, CAS maintains a computer-based inventory of published abstracts and rejected abstract manuscripts. This is an on-line operation, designed to produce management guidance tools—tools useful not only to CAS executive management, but also tools that are very useful to the CAS Library, especially in its acquisitions work. The second is a computer-produced directory of serial and non-serial titles and their ASTM Coden. The third is a computer-based project still partially in the developmental stage. This is a computer-based list of periodicals for chemistry and chemical engineering. This is an actual scheduled project with definite internal applications and with an output that will be of direct benefit to the scientific and technical information-handling community.

Before describing our computer-based inventory system, the framework for this topic should be set by referring to a few statistics which in themselves justify the system. In 1966, CAS expects to publish approximately 225,000 abstracts of published papers and patents in Chemical Abstracts. Of these, 170,000 will be abstracts of papers and government reports which originate from over 100 countries. Approximately 50 percent of this collection is published in English, while the remaining 50 percent appears in about 50 other languages. This material will come from nearly 11,000 different serials and from between 600-700 non-serial publications, such as the proceedings volumes of conferences. It will also announce the availability of

approximately 3,000 new books. To look at this another way, assume that a reasonable estimate of the world's annual output of scientific and technical articles is one or two million articles, and that a reasonable estimate of the total number of scientific and technical journals is between 35,000 and 50,000. CAS then has the problem of abstracting somewhere between 11.3 percent and 22.5 percent of the world's annual output of scientific and technical papers and monitoring somewhere between 22 percent and 32 percent of the world's scientific and technical journals.

It is apparent then from these statistics, as well as from comparable statistics for earlier years of CAS activity, that an inventory of work accomplished is an important management tool. It has been, and still is, necessary for us to have statistical data about the number of abstracts we publish, the length of the abstracts, and the number of abstracts related to various sub-disciplines of chemistry, in order to plan for future growth and to establish staff and financial requirements. CAS needs to know the source of its abstracts for abstractor remuneration purposes. It needs to know which primary journals contained the papers that were abstracted, and the relative degree of productivity of abstractable papers of these journals. CAS needs to know how many of the abstracts are prepared in-house versus those that are prepared by voluntary abstractors, and the distribution of the abstracts throughout Chemical Abstracts.

Prior to 1963, some of these statistics were available as the result of various hand counts, while others were lacking completely. In 1963 a team consisting of persons well versed in existing CAS procedures and of systems design personnel devised a computer-based inventory system which was capable of producing the needed reports from a single input operation. Although this is an on-line operation, it is not static. It is undergoing continual change. The intention is to convert from manual to machine-based input as a part of overall mechanization of CAS operation. This discussion, however, will be devoted to the existing system.

One of the factors that enabled CAS to design the present system was the ability to represent the titles of primary publications in a highly compacted form by using the American Society for Testing and Materials Coden. CAS first adopted the use of Coden for its computer-based publication, Chemical Titles. Of the 40,000 five-character unique codes that ASTM has assigned to periodical and non-periodical titles, 24,600 of these were assigned at the request of the Chemical Abstracts Service. Coden have now been assigned to all titles listed in the Chemical Abstracts List of Periodicals back to and including the 1956 edition.

Computer-Based Inventory

The computer-based inventory begins with five documents: (1) the printed issues of Chemical Abstracts in page proof format, (2) manuscripts submitted for publication in Chemical Abstracts that have been rejected, (3) a directory of periodical titles and their unique Coden, (4) the coding sheets for printed abstracts, and (5) the coding sheets for rejected abstracts. The worksheet for printed abstracts was designed in conjunction with a set of conventions which enables clerical employees to record pertinent statistical data about printed abstracts at the rate of 700-800 abstracts per day (see Figure 1). The information across the top of the form identifies the coder, the volume of Chemical Abstracts (CA), the issue, and the section which she is coding. The form is divided into four basic parts. Part 1 is for CA reference data; Part 2, the original reference data; Part 3, the name of the abstractor or the in-house abstract code; and Part 4 is the amount of space occupied in CA by the abstract and related matter, such as structures, cross references, and blank space. Under the CA reference, space is provided to record the CA column number, the location of the abstract within the column by a letter "a" through "h," and a more specific location within the letter fraction by a numeral. The language of publication of the paper is also indicated. If the abstract exceeds a certain length, this is indicated under prior lines.

For the original reference we use the Coden to designate the journal, and add the year, volume, issue, and beginning page data. Patent numbers are recorded in the column for page data while lengthy identification such as report numbers are listed in the space for other identification. The country in which the work was done and the terminal page number are also indicated.

Under the next heading is written the last name and initials of the abstractor or a four-character code that specifies an abstract produced in-house. This four-character code indicates the source of the abstract, for example, an author abstract, a translation of an author abstract, a copy abstract, or an abstract prepared by a CAS staff member. The last column is used to note by number of lines the length of the abstract.

When coding for an issue of CA has been completed, data on the completed worksheets are transcribed directly onto magnetic tape using a Mohawk Science Data Recorder, model 1101. These data are then immediately processable in the computer to produce the following reports.

Report 1 is a density report (see Figure 2). As an editorial check, it is useful to follow closely the variation of page density of

abstracts for each of the 74 sections of CA, that is, the number of abstracts in each section and the number of pages required for each section. This variation gauges the amount of editorial effort devoted to each section. These reports are produced biweekly. They show: (1) the number of abstracts of papers and patents in each section; (2) the amount of space, in columns, devoted to abstracts of papers and patents in each section; and (3) the total number of abstracts and columns for each section. Each issue report also shows these same data cumulated for all issues of the volume being inventoried.

Report 2 is the in-house abstracting report. Each CAS abstract produced in-house is coded in CA with a descriptive four-character code. Statistics for abstracts prepared by full time CAS staff become important as new services such as Chemical-Biological Activities are offered, and as the use of author abstracts changes. This report is prepared for each volume of CA. It provides a complete inventory of in-house abstracts showing the total number in each issue (see Figure 3), and their distribution by section (see Figure 4).

Report 3, a three-part report, is designed to aid library acquisition work. It is important for us to know the sources of the papers that are abstracted and to relate these data to the acquisition and processing of the journals throughout CAS.

Report 3A (see Figure 5) lists alphabetically by Coden all of the primary and secondary journals, patents, books, and one-shot publications (conference proceedings, monographic collections of papers, etc.) represented by the abstracts in a volume of CA. Opposite each Coden is the number of abstracts from the publication. This report enables CAS to determine a cost/use ratio for the documents it purchases.

Report 3B (see Figure 6) is arranged in descending sequential order by the number of abstracts each document produced. That is, it begins with the journal which contained the most papers abstracted and ends with the one which contained the least. This report emphasizes the more productive journals and serves as a guide for establishing candidate journals for express handling. For each primary journal and each one-shot publication this report shows:

(1) cumulative number of primary journals and one-shots,

(2) abstracts per title,

(3) cumulative number of abstracts of primary journals and one-shots,

(4) percent of total number of abstracts at 5-6 percent intervals,

(5) cumulative number of primary journals,

(6) cumulative number of abstracts of primary journals,

(7) percent of total number of abstracts for primary journal papers based on 5-6 percent intervals,

(8) cumulative number of one-shots,

(9) cumulative number of abstracts of one-shots, and

(10) percent of total number of abstracts for one-shots based on 5-6 percent intervals.

The tapes used to generate these volume reports are merged, and listings of journals with cumulative productivity data are printed. These cumulative lists give us a truer picture of the journals which are most productive, and these are the journals we handle on an express basis. By cumulating data from these reports for four years, and plotting the number of journals required to produce any given percentage of the total number of abstracts, an interesting curve is obtained (see Figure 7). The curve shows that 25 percent of the total number of abstracts were obtained from 50 journals; 50 percent were obtained from 250 journals; 75 percent were obtained from 1,000 journals; while 90 percent came from slightly over 3,000 journals. The remaining 10 percent of the abstracts came from an additional 2,500 journals. We monitor an additional 8,000 journals of which an average of 2,500 produce the final 10 percent of the total number of abstracts.

Report 3C (see Figure 8). The growth in many areas of chemistry and chemical engineering is gauged by monitoring other secondary journals that overlap CA coverage. We currently monitor approximately 50 such secondary publications. This report lists the primary journals that contain papers covered as copy abstracts.

Report 4 is used to generate the CA abstractors' payroll. The output is the abstractors' actual pay checks. CAS pays its abstractors by the line as printed in CA. Therefore, it is necessary to determine for each abstract the number of lines and to record this information along with the abstractor's name. The payroll is produced on a quarterly basis.

Report 5. Certain areas of chemistry produce a great many index entries per abstract. For instance, papers dealing with chemical synthesis report a great many compounds. The corresponding sections of CA are indexed directly from the original journal article. Thus for the indexers of these sections their primary document needs can be predetermined. One of the products of the inventory is the direct print-out of library call slips (see Figure 9) for those sections which have a high density of index entries per paper. This permits the library to arrange the original journal material in a manner convenient for the indexing effort.

Report 6 is a listing of document titles, with the number of abstracts from each, by subject based on CA's 74 sections. Either single sections or a combination of sections can be listed. Coverage within each of the individual sections of CA is continually watched. In addition, each year we select several sections to investigate in depth. Here we review with the section editor those subdivisions

which he feels deserve more attention, and the primary journal sources for each subdivision are investigated.

Report 7 is a listing of references to printed abstracts arranged according to bibliographical citations by primary journals. It is used to check coverage of a given journal. When merged with Report 8, which is described below, it serves as a useful tool to search for holes in the coverage. We intend to extend the usefulness of this report by inputting data about papers that have been perused but not selected; thus, any gaps in the coverage will be very apparent; and, in fact, the report will take a negative approach to the problem, i.e., it shows us what has not been entered into the CAS system.

Report 8. Another aspect of the computer-based inventory is the record of abstract manuscripts that have been rejected. Each year a number of the abstracts submitted to us for publication do not meet the criteria for publication in CA. For example, they do not report chemical information or new chemical information. Others are duplicates of abstracts that have already been published. Duplicates fall into two classes: those papers which are abstracted twice by error and those papers that are published in two or more sources. These rejected abstracts are coded on worksheets (see Figure 10), transcribed onto magnetic tape, and reports are generated which show (1) the primary journal address of the article, (2) the name of the abstractor, (3) why the paper was abstracted, that is, did the abstractor do it voluntarily or did we ask him to do it, and (4) why the abstract was not used.

CAS Coden Directory

A basic tool for all of our computer-based work which involves the titles of journals and abstracted monographs is the CAS Coden Directory. This Directory now contains an alphabetical listing of serials and nonserials by full title, each accompanied by a unique five-character Coden with a numerical sequencing or alphabetizing number (see Figure 11). The Directory has an index arranged alphabetically by Coden. The data contained in the Directory is all in upper-case type. We are now in the process of writing programs which will enable us to revise the Directory. The new Coden Directory will contain (1) the Coden, and in upper and lower case type (2) the full title, and (3) the American Standards Association title abbreviation (see Figure 12). This Directory will also contain a numerical sequencing number and will be accompanied by an index arranged alphabetically by Coden. This Coden Directory constitutes an important source of input for the periodicals list as will be explained later on.

Periodicals List

Before describing the work that we are doing on our new List of Periodicals, the history of this continuing project is worthy of mention. CAS has always paid conscientious attention to nomenclature. This applies not only to the names of chemical compounds, but also to the titles of chemical publications. In 1906 the American Chemical Society published a list of journal titles in the Proceedings of the American Chemical Society. It was entitled "A List of Abbreviations" and contained 37 journal titles and the abbreviated form for each. This list was revised, enlarged, and issued as a separate publication under the promising title List of Journals to be Abstracted by Chemical Abstracts. It preceded by several months the first issue of Chemical Abstracts and consisted of abbreviated titles for 370 journals. In 1908 the list was again revised and published as a part of the introduction to the second volume of Chemical Abstracts. It contained 475 abbreviated titles and was accompanied by three simple rules pertaining to journal title abbreviation. In 1910 the list appeared as the List of Periodicals, and the form of entry still in use today was introduced. The full title was given with those letters constituting the abbreviation set in bold face type. In addition to the title and abbreviation, the entries included frequency, volume number, price, and the publisher's address. The List continued to be issued biennially until 1922, when a second important feature was added—library source-guide data for 162 American libraries. This key to library files gave new value to the List, for it now became a useful tool for locating abstracted journals in the libraries of the United States.

The purpose of the List was two-fold. First, it was to standardize the abbreviated title form. Second, it was to furnish bibliographic data about abstracted papers by providing a library source guide. Beginning in 1936 the List was published quinquennially. In 1957 the first annual supplement to the quinquennial edition appeared.

CAS has now begun work on a completely new edition of the List of Periodicals. This new List is being designed to include much more data than did earlier editions. It is to be made available in both printed and machine-readable form (either punched cards or magnetic tape). Programs for searching the tapes will also be made available.

This catalog of periodicals related to chemistry and chemical engineering will bring into a single source much valuable bibliographic data in considerable detail that is now only partially available and from a large number of bibliographies, union lists, and individual library files. A typical entry in the List will include the full title of the publication; a translation of the title if the language is other than English, French, German, or Spanish; the ASA Z39 standard title abbreviation; the ASTM Coden; the languages of publication and

summaries; the history of the periodical with references to any former titles, current frequency, and volume number data; the publisher's address; and the price. The entries will also include the title cataloged according to American Library Association cataloging rules and the library holdings information (note that we will list holdings rather than merely an indication that a library has a current subscription to a journal).

Initially we intended to publish a 1966 edition of the Chemical Abstracts List of Periodicals with Key to Library Files. This List would have been merely an updated version of the 1961 List. The relationship between it and the 1961 List would have been identical to the relationship between the 1961 List and the 1956 List, that is, already published information about discontinued periodicals and name changes would not have carried forward into the new List. Had this course been followed, the 1966 List would have been the 14th CA List of Periodicals. To do a literature search using all of Chemical Abstracts, one would have had to use all 14 editions. As an alternative we investigated the possibility of publishing a 60-year cumulative edition of the List of Periodicals, thus bringing within one set of covers pertinent bibliographic data about, and library source guide information for, the serials listed in all previous editions.

The next logical extension of the List was to consider the inclusion of data for other serial publications related to chemistry and chemical engineering which had not been abstracted by Chemical Abstracts. A survey of Beilstein's Handbuch der Organischen Chemie revealed some 300 pre-1907 titles. Coverage of the literature of pure and theoretical chemistry from 1830 through 1940 by Chemisches Zentralblatt (CZ) was particularly outstanding. We discovered approximately 500 journals covered by CZ prior to 1940 had not been covered by CA. We also developed a list of approximately one hundred defunct nineteenth-century chemical journals to be included.

Thus the List which is proposed will include approximately 24,000 entries for journals, and 5,000 entries for monographs abstracted by Chemical Abstracts, Chemisches Zentralblatt, and covered by Beilstein, as well as the nineteenth-century chemical journals never covered by any of these services. This extension makes our List truly a comprehensive list of periodicals for chemistry and chemical engineering.

The procedure used to generate this List may actually be viewed as two separate projects: (1) gathering the pertinent bibliographic data, and (2) obtaining library holdings information from some 350-400 cooperating libraries. Our first step was to establish a file of the serial and nonserial titles. This was done by cutting and pasting the entries from all of the Lists of Periodicals, entering them into a single file, and adding to this file titles for periodicals from Chemisches Zentralblatt, Beilstein, and several other sources. Once the

file was established, the Coden was added to each main entry. This necessitated the acquisition of several hundred additional Coden from ASTM (see Figure 13).

The first publication our work generates is a checking edition for the List of Periodicals (see Figure 14). The checking edition contains the titles cataloged according to ALA rules, the Coden, the history of the serials, and a space for the cooperating libraries to write in the dates of their consecutive holdings. It is necessary to furnish the cooperating libraries with a checking edition arranged according to ALA cataloging rules, since the majority catalog their serials according to these rules. The checking edition for the CA 1961 List was produced on unit card equipment. It contained 9,000 entries in ALA form. Fortunately, these cards had been retained.

A new checking edition is now being prepared in three parts for use by the cooperating libraries. The punched card with the ALA form of entry is matched with 5" x 8" cards from the master file. If no punched cards are available, the ALA form of entry is written on the 5" x 8" card. Historical data are added and the cards passed to keypunchers to generate a new set of punched cards containing the Coden, the title in ALA form, and the history. These cards are interfiled. When the file is complete, they will be converted to tape. The resulting print-out will be used to produce camera-ready copy for printing the checking editions.

These checking editions will be sent to the cooperating libraries, and when each is returned, cards will be punched showing the Coden, the cooperating library's code number, and the library's consecutive holdings of the serial (see Figure 15). These cards will be read onto tape, sorted, and a tape containing library holdings information will be ready for future use.

There will be then one tape with the full title of a publication in upper and lower case, its official ASA Z39 abbreviation, and ASTM Coden. On another tape there will be the ASTM Coden, the ALA form of entry for the publication, and the history data.

The cards in the master record file, which contain Coden and entries for the publications as they appeared in a CA List of Periodicals, will be compared to current issues of the serials and nonserials. The entries will be edited, expanded, and updated (see Figure 16). The data on the edited master record file card will then be keyboarded on a Dura Mach - II paper tape generating typewriter for later conversion to magnetic tape. The worksheet is being designed so that the Dura typist will not have to type any of the information already in machine-readable form, i.e., she will not have to keyboard the title or abbreviated title, the history, or the ALA form of entry. These will all be retrievable by inputting the Coden. She must add needed subtitles, English translations, references to the former titles, languages of publication and summaries, the frequencies of

publication, the current volume numbers and years, prices, and the identification of the publishers. Repetitive phrases such as "Printed in," "with_____ summaries," etc., will be stored in the computer so that these do not have to be keyboarded but only coded.

Entries for discontinued periodicals, for former titles of periodicals, and for congress and symposia proceedings volumes will contain those elements of information that are pertinent to them. For instance, entries for defunct titles will not contain publisher or price data. For congress and symposia proceedings volumes, the number of the meeting, if it is one of a series, the place where it was held, the date of the meeting, the number of volumes of papers published, and the editor's name will be entered.

After the paper tape has been generated and converted to magnetic tape, this will be entered into the computer, matched with the data from the Coden directory tape, data from the tape containing the ALA form of entry, and the history information. The output will be a proof copy which will then be checked against the original edited 5" x 8" card. If any corrections need to be made, the proof copy will be edited and a corrected paper tape generated and recycled. The result of this work will be the bibliographic data tape. This tape will be merged with the library holdings tape and a master record tape generated from which a print-out will be produced for camera-ready copy for the printed version of the periodicals list.

From this master record tape, it will be possible to generate special listings of the journals by specified characteristics such as country of origin, language, price, etc., or by any combination of the elements which constitute the complete entry. Library holdings and union lists will also be obtainable (see Figure 17). The present production schedule calls for the completion of the Comprehensive List of Periodicals by the end of October 1968.

Chemical Titles

The CAS computer-based system really began in 1961 with the publication of Chemical Titles (CT), the first production-scale application in the world of computers to produce an index to technical papers. CT was the forerunner of a whole new generation of mechanized services. CT is a biweekly, computer-based publication which covers titles of chemical and chemical engineering interest taken from 691 journals. Approximately 90,000 titles were published in CT in 1965. All articles covered in CT are also covered in CA. The majority of the journals covered is received at CAS in advance form and thus the coverage is quite prompt. The titles are selected by the professional staff at CAS, translated, and then edited. Editing

of the titles includes segmentation of words to facilitate scanning of the indexes by users. The article titles, Coden (a five-letter code for the journal title) references, and authors are keypunched into cards, and the information then transferred to magnetic tape for computer processing. Check lists (corresponding to galley proofs in standard publication procedures) are printed by the computer, and the lists are proofed to insure the accuracy of the keypunched data.

Each two weeks the input data is processed by the computer and sorted into the three parts of CT: the Keyword-in-Context (KWIC) Index, the Bibliography Section (tables of contents), and the Author Index for the issue. The information is printed by the computer on graphic arts quality paper and is composed into pages mounted on special cardboard backing. The bold face journal titles are prepared on a Varityper Headliner and pasted on the pages. The camera-ready copy is then sent to the printer, where it is photographically reduced and printed on an offset press.

The KWIC index is an alphabetical arrangement of keywords selected from the article titles (see Figure 18). The keyword appears in the center of the page with its associated words preceding and following it in the same manner as they do in the title itself. Since the number of words surrounding the indexed keyword is limited by the breadth of the page, words or phrases are sometimes truncated. If the keyword appears near the beginning or ending of a title, the title will "wrap around" to fill the line space. A "wrap around" is indicated by a small cross in the text. Each keyword index entry carries a reference to the Coden, volume (or issue), and page of the journal in which the title can be found (and by which entries are listed in the bibliography section). The KWIC index provides an average of six entries to each title. The bibliography section lists the journals arranged alphabetically by Coden (see Figure 19). For each Coden, the journal title, volume, and issue are given. The titles taken from each journal are grouped together. The author index lists all authors alphabetically (see Figure 20). Each index entry provides a reference to the bibliography section.

CT is thus arranged to permit the location of information of interest via keywords, authors, or journals. It is apparent that the magnetic tapes from which the issues of CT are produced can be useful in special subject searches. During 1964, several cooperative experiments were conducted involving CT tapes. In one of the experiments, individual scientists at Eli Lilly and Company provided search terms covering their individual interests. Copies of the magnetic tapes, corresponding to each CT issue, were delivered to the company and computer searches were performed. Search results in the form of individualized listings of papers were given to the participating scientists. Each individual was allowed to change his set of search terms as his interest became better defined or as it changed. In

another experiment, searches were performed at CAS for the Olin-Mathieson Corporation. In this program the set of search terms represented the interests of a research group.

The highly successful results of these experiments and the substantial interest shown by other organizations provided the basis for the introduction in 1965 of subscriptions to CT tapes and searches. A subscriber may receive copies of the magnetic tapes or he may have his searches run by CAS in Columbus, Ohio, with the search results supplied in the form of a computer print-out. The tapes and searches are provided at the time that CT copy is sent to the printer —about a week ahead of the appearance of the CT issues. Along with the tapes are provided search programs written for an IBM 1401 or 1410. These programs are now being rewritten for the IBM System 360 computers and are expected to be available by May, 1966. The program can handle "and," "or," and "not" logic. It is also possible to specify word fragments as search terms. These aspects of the program will be explained in detail later. The search product can be any or all of the following: a bibliography of the selected titles, a KWIC index of the selected titles, or an author index based on the selected titles.

As part of the production of CT, a Coden-volume checking feature is incorporated. The check is based upon sets of valid Coden and Coden-volume relationships stored in the computer. Each time that information is input to the computer, it is tied to the appropriate bibliographic data. A computer check is made to determine the validity of these data and a list of invalid relationships is printed with each check list. In this manner errors can be detected prior to the production of camera-ready copy.

Chemical-Biological Activities

It has been said and written many times in recent years that the sheer bulk of primary literature being published makes it almost impossible to keep abreast of new information in a given field. This is particularly true in the biochemical fields. The number of abstracts published in the biochemical sections of Chemical Abstracts in 1965 is estimated at 63,000. This is 32 percent of the approximately 195,000 abstracts published for all fields. The figure is based on coverage of approximately 11,000 journals, in 50 languages, from 100 countries.

When the magnitude of this amount of information is considered, it becomes readily apparent that specialized alerting and retrieval services are needed. To help meet this need, CAS introduced Chemical-Biological Activities (CBAC) in January 1965 to provide a

rapid survey of literature dealing with the biological activity of organic chemical compounds. CBAC is a biweekly, computer-based publication issued in two volumes annually.

In the development of CBAC, a number of features were recognized as being desirable for a current-awareness publication in the field of chemical-biological activity. Among these are: promptness; adequate coverage of pertinent journals; information on the journal sources, including names and addresses of the authors; a format that combines legibility with ease of scanning; brief but adequate digests of the work reported (including descriptions of the substances employed, structural formulas, conditions under which the observations were made, and an account of the results obtained); and finally a set of indexes and a method of storing data such that the indexes can be cumulated.

The developmental work on CBAC, including the original concepts, the systems development, and the necessary computer programming, was done by the CAS Research and Development Division, and this division continues to assist as the system is modified and improved.

CBAC is a service which combines the values of an alerting tool, through printed abstracts and indexes of new developments in biochemistry, with computer-based storage and retrieval of information. As such, CBAC's function is broader than that of either a repository for information or an abstract publication alone. In the latter capacity, as has been mentioned, it is a biweekly publication. Its abstracts are known as "digests" to distinguish them from those published in Chemical Abstracts. Since a considerable portion of incoming material earmarked for CBAC is received by the Library in advance form, the coverage is quite prompt.

Considered on a broader scale, CBAC is not only a self-sustaining entity but is an integral part of a larger system for specific chemical information retrieval, the CAS Chemical Compound Registry. The Registry is based on the assignment of individual numbers (Registry Numbers) to each organic compound of known structure which enters the system. The Registry Number assigned to a compound of known structure identifies it completely—different ionic forms, isomers, and compounds containing unusual isotopic masses are given unique numbers. Use of the numbers provides a specific method of storing and retrieving information on any registered compound within the CAS system. CBAC, then, is a tool whose purpose is to make the literature of its defined fields more readily available to the scientist, both immediately and subsequently.

CBAC serves a field in which there is a great deal of industrial, academic, and governmental interest. At present, digests are prepared from articles taken from 350 English journals, 70 German, 38 Russian, 35 French, 18 Italian, 17 Japanese and 20 journals in 7

other languages. These primary journals are examined as rapidly as they are received at Chemical Abstracts Service.

To be defined as suitable for coverage in CBAC, a scientific paper must describe original work in one or more of the following areas: first, the effect (or effects) of organic compounds on the physiology or biochemistry of systems in or derived from plants, animals, or micro-organisms; second, metabolism of endogenous or exogenous organic compounds by the organisms just enumerated, or by systems derived therefrom, and third, in vitro chemical reactions, in the absence of a biological system, between organic compounds of biochemical interest.

A more detailed discussion of the criteria used in selecting articles for CBAC is appropriate. The first category, the effect of exogenous organic compounds on organisms or systems, probably applies to the largest number of papers selected for CBAC. This covers portions of the work done in almost every field of biochemistry, for example, enzymology, endocrinology, protection against irradiation, microbial biochemistry, mammalian and non-mammalian biochemistry, plant pathology, immunochemistry toxicology, pesticide and herbicide studies, and pharmacology.

The second category, dealing with metabolism, includes work in much the same areas mentioned, with the difference that here the biological organism or system is exerting an effect on a compound rather than vice versa. Since this action is usually described or characterized by alterations in chemical structure, this type of work is generally more chemical and less physiological in nature than the first category.

The third category, concerning in vitro reactions between organic compounds of biochemical interest, deals with those compounds not directly originating from a biological system in the particular experiment. It includes, for example, studies carried out using commercial enzymes in synthetic systems, as contrasted with the use of directly biological enzyme systems contained in homogenates or metachondria. Examples of other types included in this category are chemical treatment of proteins to elucidate their structures (when some biological activity is implied) or rupture of nucleotide bonds by chemical treatment in vitro, which is of interest in the mechanisms of mutations.

It need scarcely be pointed out that the interest of a given paper may overlap two or more of the three categories. For example, work reporting the effects of an antigen on a host, plus the fate of the antigen, would overlap two categories. This has no bearing on the way in which information is abstracted; all chemical-biological information is treated in context. As a result of the biological nature of the studies, the vast majority of the compounds encountered are organic. However, when pertinent, elements or inorganic compounds are dealt

with in the text. For example, excretion of potassium or other inorganic substances under the influence of a hormone or diuretic would be covered.

CBAC is divided into four sections, viz., the digest section and three indexes—the keyword-in-context (or KWIC) index, the molecular formula index, and the author index. The digest section contains the core information (see Figure 21). The digests are printed in such a way as to permit rapid scanning for subjects of interest to the reader. The names of organic compounds or classes of compounds appear at the left margin of the text lines so that their names may be read by moving the eye down the page. To the immediate left of the name for each fully defined compound, its unique Registry Number appears. The Registry Number appears only once for each compound within a given digest.

In addition, each digest contains some words which are highlighted by being printed entirely in capital letters. These are terms which, in addition to the organic names, are felt to be of interest in rapid scanning. They include, for example, names of elements and inorganic compounds (as these do not necessarily appear at the left margin), diseases and other pathological conditions, names of microorganisms (including viruses), names of organs or other anatomical compounds (mitochondria, liver), and words which indicate a function of interest in the digest (for example, oxidation or diuretic effect).

As an additional aid in scanning, structural formulas are given in the digests. These are selected by the chemists who edit the digests. The selection is based on providing a clue to the important type (or types) of organic compounds dealt with in the digests. For example, if a number of phenol analogs are discussed in a digest, only one example of a phenolic structure would be selected for illustration. A "large" structural formula is given only once in an issue and subsequent occurrences of the compound represented are cross-referred to the first appearance within the issue.

All papers selected from a given journal are printed under a bold face group heading which gives the journal name, volume, issue number, and year. Journals are listed alphabetically by Coden, rather than by full title. Each digest is headed by the title of the original paper (if in a language other than English, the title is translated). Following the title appears the complete bibliographic reference, including Coden notation, volume number, issue number, year, and inclusive page numbers. Names of authors are then given, by last name and initials, in the order given in the journal. Finally, the research site is given, when known.

All digests within a volume of CBAC (13 issues) are numbered consecutively. Each sentence in the digest is separated from the preceding sentence by a blank space. All sentences except the first are given subnumbers. The numbers and subnumbers are used in the

KWIC and molecular formula indexes to give a precise reference to that portion of the digest where the indexed entry appears. Keywords are selected by the chemist as he prepares the digest. As its full name implies, the KWIC index (see Figure 22) presents keywords in the context of the sentence within which they appear. Thus, the user is not limited to the information contained in the single indexed word ("actinomycin," for example, may be indexed several times) but he can, within the limits of a phrase, obtain an idea of the context in which the word is being used. The indexed entity appears in all-capital type in the center of the page, with its associated words preceding and following it in the same manner as they do in the digest itself. All index entries refer the user to the digest number or subnumber from which they were generated. The KWIC index provides an average of 30 entries to each digest.

The molecular formula index presents an ordered listing of the molecular composition of organic compounds to which Registry Numbers have been assigned (see Figure 23). To the right of the formula appears the compound's Registry Number; to the left the number (or numbers) of the digest (or digests) in which the compound is mentioned. The reference is given to the digest number (or subnumbers) in which the compound first appears; subsequent listings of the compound within the same digest are not indexed. The molecular formula and Registry Number do not appear opposite each digest reference, but only opposite the first reference; subsequent references are opposite a blank space. Molecular formulas are listed in order of increasing numbers of carbon atoms, and each formula is arranged according to a modified Hill system—carbon first, hydrogen second, and then other elements alphabetically by atomic symbols. (Note that in this system, CCl_4, which has no H atoms, will be listed before $CHCl_3$ or pure hydrocarbons such as C_2H4.) However, the molecular formulas of metallic salts, amine salts, and addition compounds are listed immediately following the molecular formulas of the parent substances. This is done in order to eliminate the separation that would otherwise occur between parent and derivative. Thus, following the molecular formula of a carboxylic acid, references to its sodium salt would appear. Even though no actual reference to the parent compound may appear in the given issue, its molecular formula will be given in the index to facilitate finding the molecular formulas of derivatives which are referenced.

Finally, the author index presents an alphabetical listing of all author names in the issue, each name being matched with the corresponding digest number (or numbers). The three indexes are cumulated and issued every six months (corresponding to volume indexes).

The steps involved in producing CBAC are briefly as follows (see Figure 24). Journals which are covered in CBAC are scanned in the Assignment Department. A biochemist reviews each article in

depth to determine if it deals with one or more of the subject criteria used in the selection of CBAC oriented material. Digests of the articles selected are prepared by chemists at CAS. All articles covered by CBAC are also covered in CA. The digests, which are on standard forms, are sent to our Formula Indexing Department where registry sheets (containing structural formulas) are prepared for each of the organic compounds with defined structures. The digest information is then keypunched into cards and entered on magnetic tape by means of the computer system. Paralleling the processing of digests, the registry sheet information is being processed in the CAS Registry System, where Registry Numbers are assigned to the compounds by the computer. These numbers are stored on tape for merging with the corresponding compounds in the text of the digest. Computer worksheets for each of the digests are edited and any corrections reentered onto the computer tape. The information on tape pertaining to the digest section is rearranged by the computer to provide the KWIC index, the molecular formula index, and the author index. Pages of the digest section and the indexes printed by the computer are pasted on cardboard sheets and photographed. The film is used to prepare plates for offset printing of the CBAC issues.

CBAC is constantly being evaluated for possible improvements. A case in point is an additional index included in the cumulated indexes of Volume 2 (end of 1965). This is a Registry Number and Registry Number Cross-Reference Index that will not only permit a search of the digests for individual compounds, but also for compounds whose structural features are similar. For example, an amine will be cross-referred to its acid salts by Registry Numbers and vice versa.

Starting in January of 1966, CAS began to provide, on a subscription basis, the magnetic tape records of CBAC, to be searched at the subscriber's location or at CAS. The tapes include all of the information contained in the digest section, as well as the molecular formulas of the registered compounds. This permits searching of the tapes for chemical compounds by names and by Registry Numbers and also searching for syllables, suffixes, prefixes, words, or phrases according to a given interest profile. Thus, it will be possible to conduct a search for chemical-biological activity correlations. The output of the search will be bibliographic references (CBAC digest number, original journal by Coden, volume, issue, and page, as well as titles and authors.) A KWIC index of the titles cited or a list of the authors will also be available, if desired.

One of the features that will permit great flexibility in searching for compounds is the ability to use Registry Numbers as search terms. This method will permit location of registered compounds in digests, regardless of the chemical names used to describe the compounds. The search programs permit "and," "or," and "not" logic.

Weighting factors are available in the search programs to increase the relevancy of the data recovered.

The present search programs are designed for use with an IBM 1401 or compatible computer. However, the search is being reprogrammed in a higher language, and it is expected that it will be complete by May 1966. The new programs will be oriented toward the IBM system 360 family of computers and should permit reprogramming to utilize other computer configurations. The same Coden and Coden-volume computer checks as discussed in conjunction with CT are applied during the production of CBAC.

CAS Chemical Compound Registry

It was mentioned previously, in the discussion of CBAC, that Registry Numbers provide a specific method of storing and retrieving information on registered compounds within the CAS system. The Chemical Compound Registry is the heart of the CAS computer-based system. The Registry files must contain a detailed description of each compound. At the minimum, the computer record of each compound includes the full detail of the conventional structural diagram. Chemical nomenclature does not form an adequate basis for the computer record. The Registry file also supplies the basis for substructure searches—searches for fragments of molecules. Again, nomenclature does not provide enough details for direct substructure searching in depth. It is important that the computer files include no undetected synonomy, which if uncontrolled would increase the files to unmanageable proportions.

A tabular description of the structural diagram of compounds is used in the Registry. The description, which is called a connection table, lists each of the atoms in the molecule atom-by-atom and bond-by-bond. Programs have been developed to manipulate this tabular description so as to convert the table to a unique, unambiguous form. The form is not quickly intelligible to the chemist, but it is easily handled by the computing system. Thus, the Registry file of structures consists of an ordered list of unique, unambiguous descriptions of the individual compounds. The registration process amounts to an ordering and merging by computer. The purpose is to determine whether a particular structure has already been stored in the system. If a compound has been registered previously, the same Registry Number is supplied as output; but if a compound has not been registered, a new Registry Number is provided.

Future Computer-Based Operations

CAS has plans in preliminary design state for several automated library systems capable of handling a large input of documents on a very short time-span basis. This effort is but a part of the overall work being done to put CAS on a computer base.

The master record tape produced for the publication of the Comprehensive List of Periodicals will serve as the basis for automating many of our existing library systems (see Figure 25). Existing serials inventory files, source files, disposition data files, and coverage data files are to be converted into machine-readable form and these data are to be organized in disk storage. Once the data are in machine-readable form, it will be possible to make many uses of them. The library applications include automated serials check-in and rapid retrieval of routing or disposition information. At CAS the latter is extremely important. Misrouted serials can result in duplication of abstracting or in papers not being abstracted at all. It now takes between six and nine months to train fully a clerical employee to record and route incoming new materials with minimum supervision. Under our new system this training period can be reduced by the length of time it takes to train the clerks to inquire of the computer what it is that they are supposed to do with the issues at hand.

We will be able to provide printed catalogs of the serial holdings and receipts with coverage data. We also intend to reproduce our existing dictionary card catalog in machine-readable form and to produce printed book catalogs. This process will also generate the book announcements for publication in Chemical Abstracts.

PRINTED ABSTRACTS

CODER	CA VOL.	CA ISSUE	CA SEC.	X 180 + = PAGES LINES FRAC. PG. LINES TOTAL

CA REFERENCE				ORIGINAL REFERENCE								ABSTRACTOR NAME OR 4 LETTER CODE	NUMBER OF LINES			
COLUMN	FRACTION SUB PAR.	LANGUAGE	PRIOR LINE	CODEN	YEAR	VOLUME	ISSUE	PAGE, PATENT, OR ABSTRACT NUMBER	OTHER IDENTIFICATION	COUNTRY OF WORK	ENDING PAGE NO.		TEXT	STRUCTURE	X-REF	BLANK
5	1 1	2	2	4 1	2	4	6	7	21	3	3	15	3	3	3	3

Form 60-133 Revised 2/17/66 CHEMICAL ABSTRACTS SERVICE

Figure 1
Worksheet for Coding of Printed Abstracts

VOLUME 64 STATISTICAL REPORT

SEC	ISSUE NUMBER 4 PAPERS ABST	PAPERS COLUMNS	PATENTS ABST	PATENTS COLUMNS	TOTALS ABST	TOTALS COLUMNS	ISSUES 1 THRU 4 PAPERS ABST	PAPERS COLUMNS	PATENTS ABST	PATENTS COLUMNS	TOTALS ABST	TOTALS COLUMNS
1	36	2.57		.00	36	2.57	141	10.74		.00	141	10.74
2	144	24.83	3	.37	147	24.90	659	99.88	33	3.89	692	103.77
3	207	32.51		.00	207	32.51	646	99.71	1	.12	647	99.83
4	76	14.58	4	.53	80	15.11	244	42.06	14	2.59	258	44.65
5	50	10.12	6	1.20	56	11.32	166	28.12	19	3.05	185	31.17
6	94	16.64		.00	94	16.64	376	60.91		.00	376	60.91
7	40	6.34		.00	40	6.34	155	23.09		.00	155	23.09
8	180	29.28	4	.47	184	29.75	507	75.39	14	2.38	521	77.77
9	260	43.26	43	7.21	303	50.47	917	146.45	113	19.96	1030	166.41
10	388	64.98	1	.12	389	65.10	1160	182.80	11	1.86	1171	184.66
11	40	6.12	33	6.45	73	12.57	125	21.75	135	32.97	260	54.72
12	151	22.58		.00	151	22.58	1003	143.13	1	.17	1004	143.30
13	125	20.56	18	3.15	143	23.71	603	86.94	70	11.74	673	98.68
14	129	21.35		.00	129	21.35	416	70.47		.00	416	70.47
15	140	22.05	50	7.44	190	29.49	708	125.49	250	39.91	958	165.40
	2060	337.47*	162	26.94*	2222	364.41*	7826	1,216.93*	661	118.64*	6487	1,335.57*
16	186	23.73	86	11.75	272	35.48	607	80.19	170	22.12	777	102.31
17	29	5.23	76	11.96	105	17.19	116	18.88	238	37.96	354	56.84
18	58	9.92	23	3.78	81	13.70	189	30.63	69	10.09	258	40.72
19	150	27.44	46	8.72	196	36.16	590	107.68	143	24.64	733	132.32
20	147	23.18	59	11.05	206	34.23	544	89.67	147	25.67	691	115.34
21	104	18.88	42	7.36	146	26.24	336	58.17	119	20.45	455	78.62
22	18	2.04	15	1.97	33	4.01	96	13.91	31	4.01	127	17.92
23	40	6.02	9	1.33	49	7.35	150	24.18	28	3.85	178	28.03
24	44	5.23	7	.87	51	6.10	208	28.55	35	4.88	243	33.43
25	129	19.94		.00	129	19.94	774	130.33		.00	774	130.33
26	44	7.02	8	.77	52	7.79	122	20.25	23	2.92	145	23.17
27	80	14.34	68	12.77	148	27.11	279	47.98	270	50.99	549	98.97
28	28	6.43	8	1.38	36	7.81	98	20.66	36	7.28	134	27.94
29	17	2.08	8	1.39	25	3.47	59	7.20	16	2.85	75	10.05
30	87	13.44	34	4.52	121	17.96	247	34.18	116	17.90	363	52.08
	1161	184.92*	489	79.62*	1650	264.54*	4415	712.46*	1441	235.61*	5856	948.07*
31	6	.30		.00	6	.30	23	1.70	4	1.46	27	3.16
32	180	39.42		.00	180	39.42	680	142.16		.00	680	142.16
33	57	14.22	105	18.96	162	33.18	230	49.52	439	80.56	669	130.08
34	19	10.58	17	4.62	36	15.20	87	29.33	89	18.85	176	48.18
35	63	17.57	84	25.36	147	42.93	305	87.06	350	91.89	655	178.95
36	14	4.63	17	6.55	31	11.18	84	22.90	74	20.14	158	43.04
37	64	17.95	35	11.86	99	29.81	266	90.35	265	85.25	531	175.60
38	110	41.21	70	26.41	180	67.62	285	102.13	287	101.60	572	203.73
39	40	10.43	21	4.13	61	14.56	177	41.04	103	21.78	280	62.82
40	29	10.53	1	.17	30	10.70	108	38.38	13	3.80	121	42.18
41	12	6.61	5	3.43	17	10.04	54	21.86	14	7.88	68	29.74
42	27	7.71	39	17.60	66	25.31	82	26.37	95	41.08	177	67.45
43	33	9.30	8	1.60	41	10.90	161	45.02	39	7.22	200	52.24
44	35	10.47	8	3.30	43	13.77	112	33.84	37	14.05	149	47.89
	689	200.93*	410	123.99*	1099	324.92*	2654	731.66*	1809	495.56*	4463	1,227.22*

Figure 2
Example of Density Report of Chemical Abstracts

IN HOUSE ABSTRACTORS BY CODE FOR VOL 63 02/21/66 PAGE 01

CODE	NUMBER OF ABSTRACTS	NUMBER OF PATENTS
BBJN	350	
BCJD	1	
BCJG	1	
BCJN	263	
BCPC	1	
BDJF	3	
BDJL	12	
BDJN	4	
BDPF	902	902
BDPL	30	30
BDPN	88	88
BFJN	323	
BGJF	51	
BGJM	1	
BGJN	721	
BGJR	1	
BGJV	4	
BGPV	1	1
BHJN	166	
BKJG	153	
BKJN	195	
BLJB	75	
BLJC	6	
BLJD	66	
BLJF	5	
BLJH	1	
BLJM	1	
BLJN	5	
BLJR	856	
BLJS	1	
BLJT	1	
BLJV	1	
BLJW	21	
BLTJ	1	
BMJG	2	
BMJR	240	
BMJW	18	
BMJX	1	
BMRJ	1	
BNJR	1	
BPJF	15	
BPJN	91	
BPPF	3	3
BSJG	8	2
BSJN	36	
BSPG	37	37
BTJF	103	
BTJG	49	
BTJH	1	
BTJL	5	
BTJN	16	

Figure 3
Example of Computer Printout Inventory of CA Abstracts by In-House Abstractor

IN HOUSE ABSTRACTORS BY CODE BY SECTION VOL 63 02/21/66 PAGE 01

CODE	NUMBER OF ABSTRACTS	NUMBER OF PATENTS
	SECTION 1	
GLJR	1	
RCFQ	2	
RCTT	5	
RCVD	1	
RCVQ	1	
SFTT	1	
SGTT	5	
SHTT	1	
SNDC	1	
SNTT	52	
SXTT	1	
SZTT	2	
TCNG	1	
TCSL	1	
VNJZ	7	
SECTION TOTAL	82	
	SECTION 2	
BOPN	1	1
BGJV	1	
BLJB	2	
BLJC	2	
BLJD	2	
BLJR	10	
BLJS	1	
BLTJ	1	
BMJR	23	
BSPG	1	1
BTJQ	2	
BWJY	1	
BYJY	53	
GBJR	6	
GHPF	1	1
GKJY	2	
GLJR	13	
GMJR	3	
GPJR	3	
GPJW	3	
GRJF	1	
GRJG	2	
GZJR	7	
HSJR	1	
MCCR	1	
MCRK	1	

Figure 4
Example of Inventory of CA Abstracts by In-House Abstractors, Arranged by Section of CA

CHEMICAL ABSTRACTS SERVICE

CODEN BASED INVENTORY REPORT NUMBER 1 CA VOLUME 63
SOURCE OF ABSTRACTS

CODEN	NUMBER OF ABSTRACTS					PAGE 1
AAAH	9					
AAAM	1					
AABC	8	7	BY CA	1	BY OTHER ABSTRACT JOURNALS	
AABF	2					
AABI	33	29	BY CA	4	BY OTHER ABSTRACT JOURNALS	
AABN	9					
AABQ	7					
*AABS	3	1	CREDITED TO RUSSIAN JOURNALS			
AACH	79					
AACI	7					
AACR	6					
AAFA	3	1	BY CA	2	BY OTHER ABSTRACT JOURNALS	
AAFB	5					
AAFP	13					
AAFR	3					
AAGE	18					
AAGP	4					
AAGS	5					
AAGW	9					
AAHO	1					
AAKG	1					
AALS	1					
AALT	2	NONE	BY CA			
AAML	38					
AAMM	11					
AANE	6					
AANI	1					
AANL	41	37	BY CA	4	BY OTHER ABSTRACT JOURNALS	
AAOS	1					
AAPL	2					
AAQA	33					
AARE	2					
AASC	3					
AASH	1					
AATF	9	1	BY CA	8	BY OTHER ABSTRACT JOURNALS	
AATR	1					
AATS	3	1	BY CA	2	BY OTHER ABSTRACT JOURNALS	
AAUP	1					
ABAB	19					
ABAH	9					
ABAI	1					
ABBI	184					
ABBS	3					
ABCH	102	100	BY CA	2	BY OTHER ABSTRACT JOURNALS	
ABCL	18					
ABCO	6					
ABCZ	1					
ABEB	3					
ABEG	5					
ABEM	4	1	BY CA	3	BY OTHER ABSTRACT JOURNALS	

* INDICATES SECONDARY JOURNALS

Figure 5
Example of Inventory of CA Abstracts by Source in Coden Order

CODEN BASED INVENTORY REPORT NUMBER TWO CA VOLUME 63 PAPER ABSTRACTS 65859

CUMULATIVE NUMBER OF JOURNALS	CODEN	ABSTRACTS PER CODEN	CUMULATIVE NUMBER OF ABSTRACTS	PERCENT OF TOTAL	CUMULATIVE NUMBER OF PRIMARY JOURNALS	CUMULATIVE NUMBER OF PRIMARY ABSTRACTS	PERCENT OF TOTAL	CUMULATIVE NUMBER OF ONE SHOT JOURNALS	CUMULATIVE NUMBER OF ONE SHOT ABSTRACTS	PERCENT OF TOTAL
1	DIAS	2073	2073		1	2073				
2	CORE	926 *	2999		2	2999				
3	NATU	918 *	3917		3	3917				
4	JACS	799	4716	5.5	4	4716	(5.5)			0.0
5	XAER **	789	5505		5	5505				
6	JOCE	767	6272		6	6272				
7	BBAC	747 *	7019		7	7019				
8	OANK	717 *	7736		8	7736				
9	JCSO	560	8296		9	8296				
10	BIJO	439	8735	10.2	10	8735	(10.2)			0.0
11	JPCH	437	9172		11	9172				
12	PHLT	432 *	9604		12	9604				
13	JBCH	422	10026		13	10026				
14	PRVA	394	10420		14	10420				
15	ZOKH	388	10808		15	10808				
16	FTVT	387 *	11195		16	11195				
17	XCTA **	382 *	11577		17	11577				
18	0SCF	380	11957		18	11957				
19	TELE	365	12322		19	12322				
20	XRDR	344 *	12666		20	12666				
21	NAAC	340 *	13006	15.1	21	13006	(15.1)			0.0
22	PSEB	339 *	13345		22	13345				
23	KGKZ	313 *	13658		23	13658				
24	ZNUK	309	13967		24	13967				
25	CCOM	307	14274		25	14274				
26	ANCH	306	14580		26	14580				
27	IASK	305 *	14885		27	14885				
28	BCSJ	304	15189		28	15189				
29	ZFKH	303 *	15492		29	15492				
30	ZPKH	301 *	15793		30	15793				
31	PHSS	301 *	16094		31	16094				
32	PRLT	299	16393		32	16393				
33	CCCC	299	16692		33	16692				
34	NUPH	297	16989		34	16989				
35	JAPI	287 *	17276	20.1	35	17276	(20.1)			0.0
36	PRVB	285	17561		36	17561				
37	AECA	281 *	17842		37	17842				
38	CHBE	280	18122		38	18122				
39	JPMS	273	18395		39	18395				
40	CJCH	267	18662		40	18662				
41	VMSO	260 *	18922		41	18922				
42	JUPS	254 *	19176		42	19176				
43	NATW	247 *	19423		43	19423				
44	CRSB	247 *	19670		44	19670				
45	ZORK	246	19916		45	19916				
46	TAIM	244 *	20160		46	20160				
47	NUCI	244	20404		47	20404				
48	INOC	242	20646		48	20646				

* DENOTES SOME OR ALL ARE COPY ABSTRACTS ** DENOTES U.S. GOV'T PRIMARY JOURNAL OR R AND D REPORTS

Figure 6
Example of Inventory of CA Abstracts

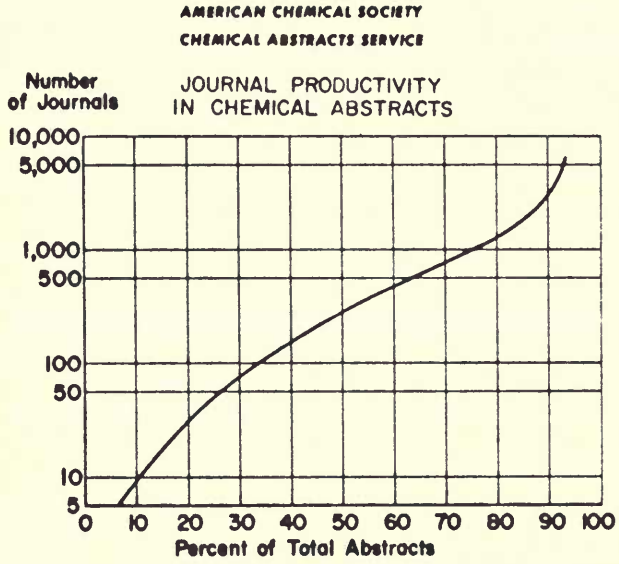


Figure 7
Four-Year Cumulation or Semi-Log Paper of Journal Productivity of Abstracts

CODEN	YR	VOL	ISS	PAGE	TITLE		PRIMARY JOURNAL CODEN	YR	VOL	ISS	PAGE	
TOTAL UNIDENTIFIED PRIMARY JOURNALS 1001												
ANAIS DA ACADEMIA BRASILEIRA DE CIENCIAS												
AABC	1962	34		181	199	$01	CHZE	1964		4	389	199 $02
	TOTAL	1										
ANNALS OF APPLIED BIOLOGY												
AABI	1963	52		327		$01	CHZE	1965			1468	$02
AABI	1964	53		59		$01	CHZE	1964		4	1230	$02
AABI	1963	52		243		$01	CHZE	1965			2337	$02
AABI	1964	53		291	309	$01	CHZE	1965		1	1269	$02
	TOTAL	4										
ANALE, INSTITUTUL CENTRAL DE CERCETARI AGRICOLE, CONSILIUL SUPERIOR AL AGRICULTURII, SERIA C.* FIZIOLOGIE, GENETICA, AMELIORARE, PROTECTIA PLANTELOR SI TEHNOLOGIE AGRICOLA.												
AAFA	1961	29		235		$01	CHZE	1964		4	2069	$02
AAFA	1961	29		9		$01	CHZE	1964		4	2048	$02
	TOTAL	2										
ARCHIVIO ,,DE VEECHI,, PER L/ANATOMIA PATOLOGICA E LA MEDICINA CLINICA												
AALT	1961	35	3	729	740	$01	BIAB	1965	46		894	$02
AALT	1961	36	2	345	382	$01	BIAB	1965	46		582	$02
	TOTAL	2										
ATTI DELLA ACCADEMIA NAZIONALE DEI LINCEI. RENDICONTI, CLASSE DI SCIENZE FISICHE, MATEMATICHE E NATURALI												
AANL	1961	30		437		$01	CHZE	1964		4	598	$02
AANL	1962	32		567		$01	CHZE	1965			1697	$02
AANL	1962	96		805		$01	CHZE	1964		4	607	$02
AANL	1962		96	517	533	$01	CHZE	1964		4	127	$02
	TOTAL	4										
ATTI DELLA ACCADEMIA DELLE SCIENZE DI TORINO. CLASSE DI SCIENZE FISICHE, MATEMATICHE E NATURALI												
AATF	1963	97		258	264	$01	CHZE	1965		1	637	$02
AATF	1962	96		89	096	$01	CHZE	1964		2	1797	$02
AATF	1962	96		903		$01	CHZE	1964		4	259	$02
AATF	1962	96		534		$01	CHZE	1964		4	688	$02
AATF	1963	97		757	765	$01	CHZE	1965		1	648	$02
AATF	1963	97		635	648	$01	CHZE	1965		1	713	$02

Figure 8
Example of Print-out Report of Inventory of Copy Abstracts Used in CA, Arranged by Primary Journal

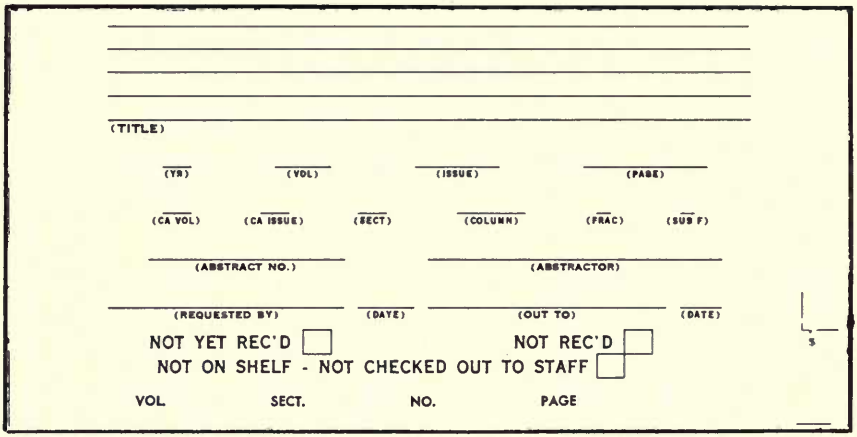


Figure 9
Library Call Slips

KILLED MANUSCRIPTS | CA VOL.

ORIGINAL REFERENCE							SOURCE OF ABSTRACT					REASON FOR KILLING				ABSTRACTOR	CREDIT	
CODEN		YEAR	VOLUME	ISSUE	PAGE, PATENT, OR ABSTRACT NUMBER	OTHER IDENTIFICATION IF ITEM CANNOT BE CATEGORIZED IN PRECEDING FIELDS	ASSIGN	WHOLE JOURNAL	COPY ABSTRACT	PROF. ABSTR.	AUTHOR ABSTR.	NOT SUITABLE	PAGE DUP.	SUB. DUP.	REASSIGN	ONLY IF CREDIT	$	¢
4	1	2	4	6	7	30	A	W	C	P	X	N	P	S	R			

Figure 10
Worksheet for Coding Abstracts

CAS CODEN DIRECTORY. TITLE SEQUENCE PAGE 1

CODEN	TITLE	SEQ. NO.
AGHY	ABHANDLUNGEN AUS DEM GESAMTGEBIETE DER HYGIENE.	000365 01
AMEH	ABHANDLUNGEN AUS DEM INSTITUT FUR METALLHUETTENWESEN UND ELEKTROMETALLURGIE DER TECHNISCHEN HOCHSCHULE. AACHEN.	000425 01 000425 02
APWP	ABHANDLUNGEN DER PREUSSISCHEN AKADEMIE DER WISSENSCHAFTEN. PHYSIKALISCH-MATHEMATISCHE KLASSE	000570 01 000570 02
ABMB	A.B.M. BOLETIN (ASSOCIACO BRASILEIRA DE METAID (SAO PAULO)) A.B.M. BOL. ASSOC. BRASIL. METAIS (SAO PAULO)	000700T01 000700T01
AVCM	ABRASIVE AND CLEANING METHODS.	000730 01
APKR	ABRIDGED SCIENTIFIC PUBLICATIONS FROM THE KODAK RESEARCH LABORATORIES.	000765 01 000765 02
ATPM	ABSTRACTS OF SCIENTIFIC AND TECHNICAL PUBLICATIONS FROM THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY.	001095 01 001095 02
ARBB	ACADEMIA REPUBLICII POPULARE ROMINE. INSTITUTUL DE BIOCHIMIE. STUDII SI CERCETARI DE BIOCHIMIE	001960 01 001960 02
ACCF	ACCION FARMACEUTICA	002690 01
AQMR	ACQUISITIONS MEDICALES RECENTES. JOURNEES MEDICALES ANNUELLES DE BROUSSAIS - LA CHARITE.	002930 01 002930 02
ABIM	ACTA BIOLOGICA ET MEDICA GERMAICA. SUPPLEMENTBAND	003630 01
ACUJ	ACTA ET COMMENTATIONES IMPERIALIS UNIVERSITATIS JURIEVENSIS (DORPATENSIS)	004045 01 004045 02
ATUA	ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS (DORPATENSIS) A.* MATHEMATICA. PHYSICA. MEDICA	004050 01 004050 02
ATUC	ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS (DORPATENSIS) C.* ANNALES	004055 01 004055 02
ALSM	ACTA LEIOENSIA. SCHOLAE MEDICINAE TROPICAE	004850 01
ALHM	ACTA LITTERARUM AC SCIENTIARUM REGIAE UNIVERSITATIS HUNGARICAE FRANCISCO-JOSEPHINAE. SECTIO MEDICORUM	004890 01 004890 02
AMTO	ACTA NIPPONICA MEDICINAE TROPICALIS.	005770 01
APAI	ACTA PHYSIOLOGICA ACADEMIAE SCIENTIARUM HUNGARICAE. SUPPLEMENTUM	006530 01 006530 02
APYP	ACTA PHYSIOLOGICA POLONICA LITTERAE SOCIETATIS PHYSIOLOGORUM POLONORUM ACTA PHYSIOL. POLON.	006580T01 006580 02 006580T01
APXG	ACTA PHYTOTAXONOMICA ET GEOBOTANICA.	006670 01
APQP	ACTA PSIQUIATRICA Y PSICOLOGICA DE AMERICA LATINA	007010 01

Figure 11
Sample Page of Title Sequence of CAS Coden Directory

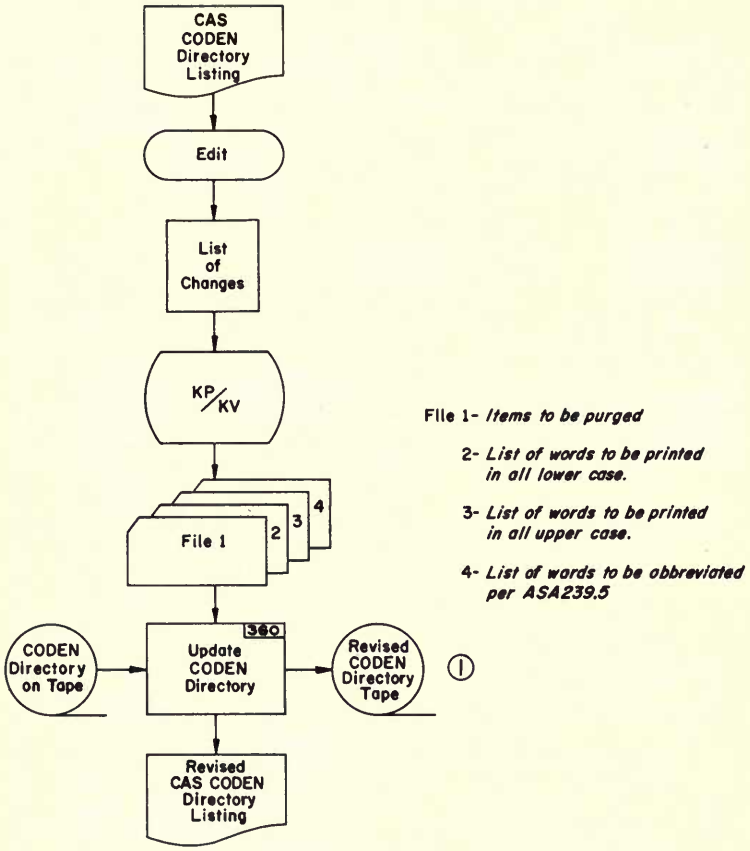


Figure 12
Flow Chart for Revision of Present CAS Coden Directory

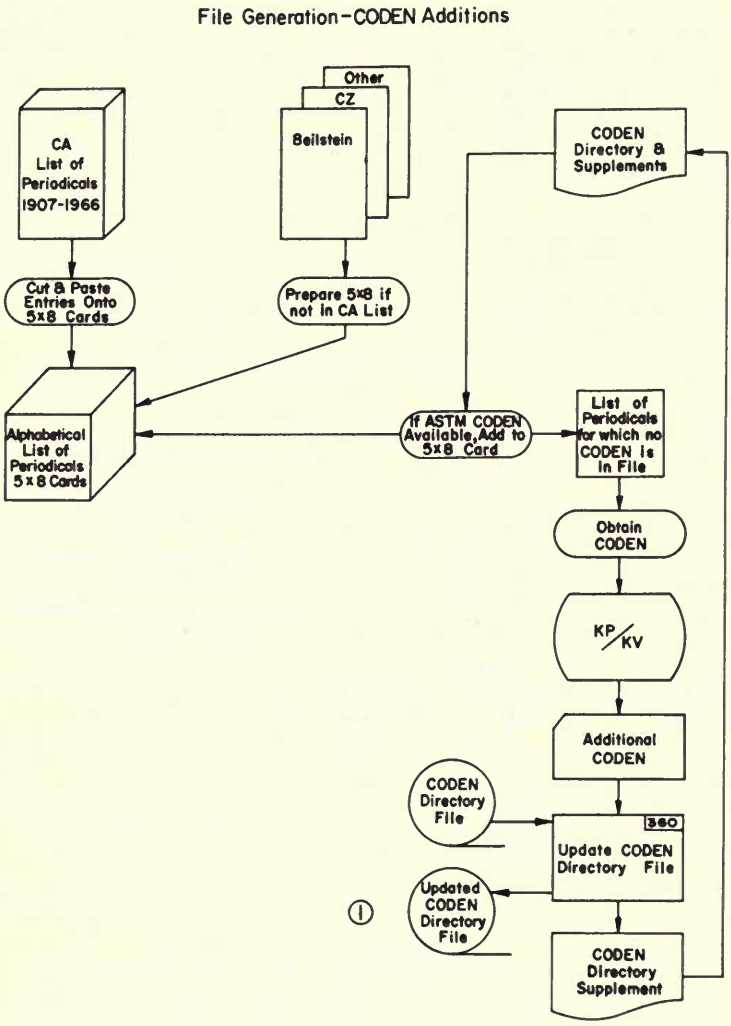


Figure 13
Flow Chart for Preparation of Master File of Periodicals with Coden

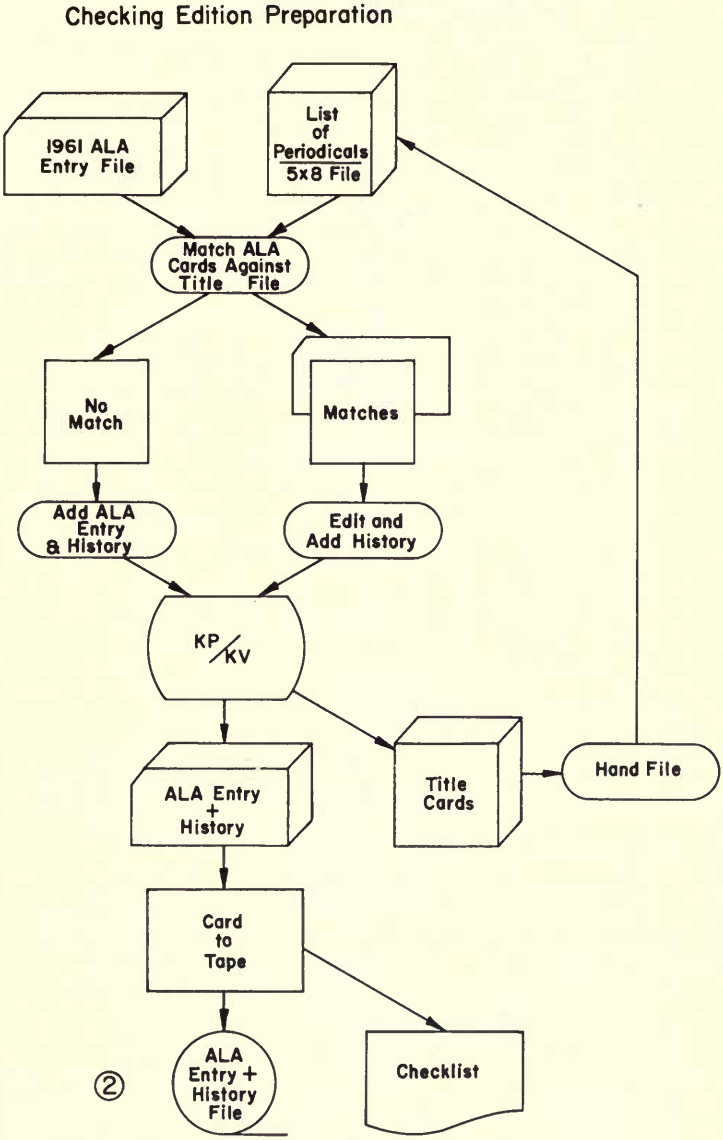


Figure 14
Flow Chart for Preparation of Checking Edition of List of Periodicals

Library Holdings Tabulation

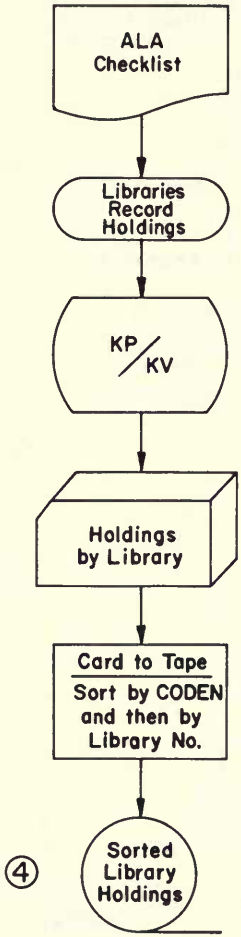


Figure 15
Flow Chart for Adding Library
Holdings of Periodicals to
Computer Record

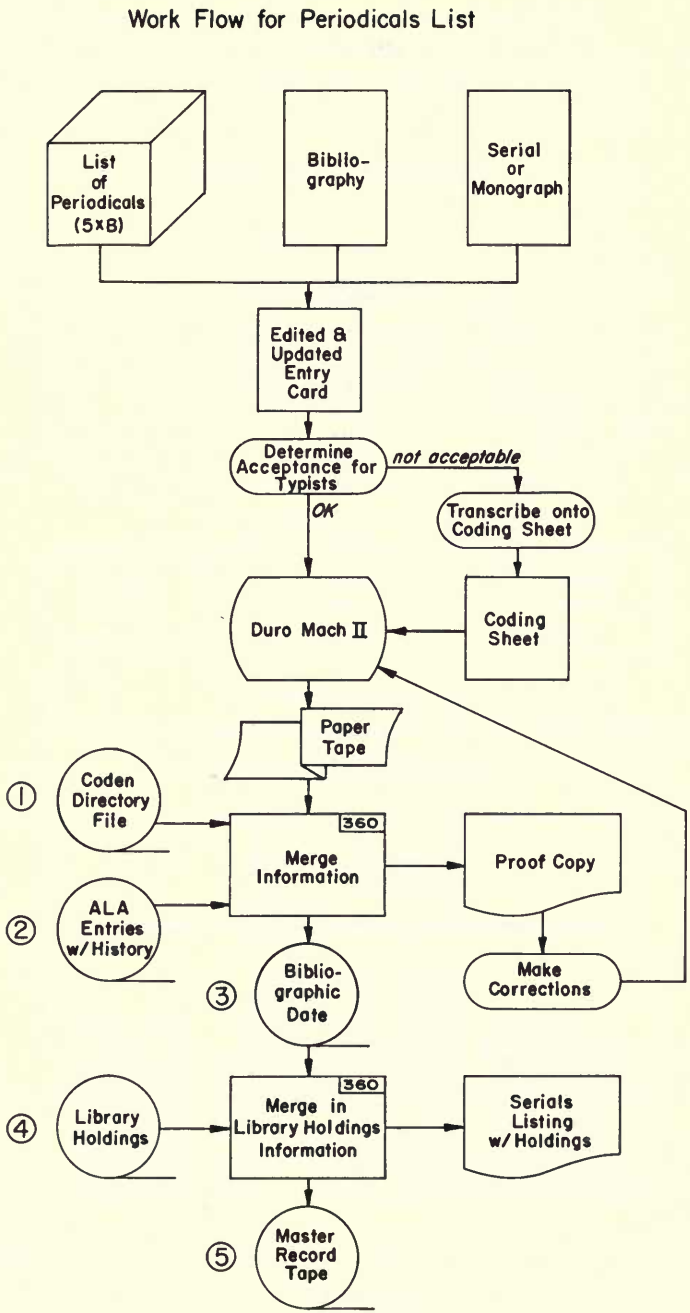


Figure 16
Flow Chart for the Preparation on Tape of the Final Complete List of Periodicals

Prepare Specialized Listings

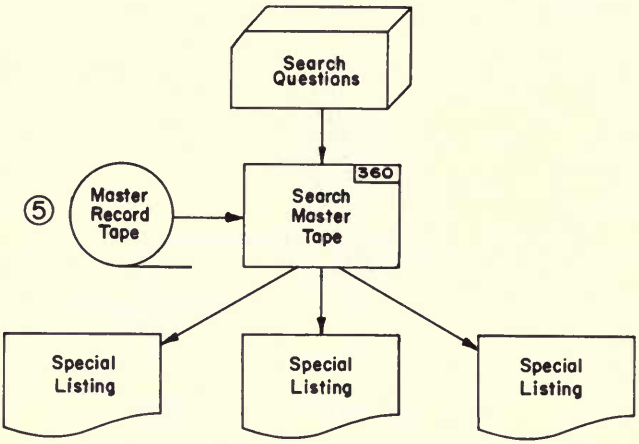


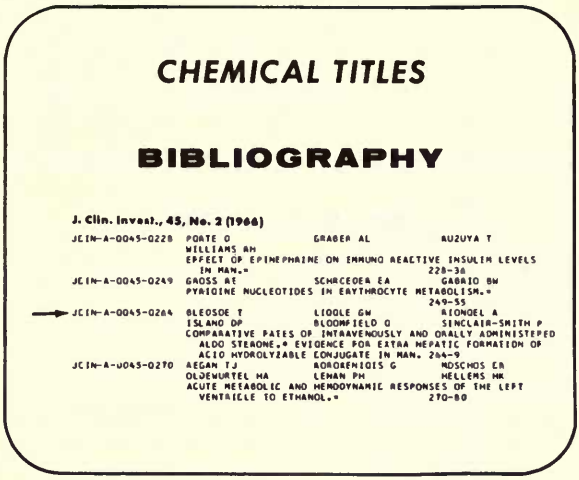
Figure 17
Flow Chart for the Preparation of Special Lists of Periodicals

CHEMICAL TITLES

KEYWORD-IN-CONTEXT INDEX

WITH AROMATIC	ALDEHYDES, AND THE GEOMETRICAL	JSOO-A-1966-0283
WITH ALIPHATIC	ALDEHYDES, MALONIC ESTERS, AND	BSCF-A-1968-3596
OF SETTING FREE THE	ALDEHYDIC GROUP IN PYRIDOXAL.=	JUAR-A-0008-0129
OF ALDOSES AND THEIR	ALDEHYDRO DERIVATIVES WITH COMPOUN	JPMS-A-0055-0028
ESTER OF THE DIELS-	ALDER ADDUCT OF PETRO SELINIC ACID	JAOC-A-0043-0033
CYCLO PENTADIENO+DIELS-	ALDER REACTIONS WITH TETRA CHLORO	TELE-A-1966-0201
AND ORALLY ADMINISTERED	ALDO STERONE.+ EVIDENCE FOR EXTRA	JCIN-A-0045-0264 ←
RIBOSE-5-PHOSPHATE	ALDOLASE IN THE SYNTHESIS OF DEOXY	BBRC-A-0022-0062
OF METHYL DOPA I	ALDOMETI ON CARDIOVASCULAR RESPONS	AIPT-A-0159-0215
TIVES + CONDENSATION OF	ALDOSES AND THEIR ALDEHYDRO DERIVA	JPMS-A-0055-0028
SYNTHESIS OF 2,2-	ALDOSYLIDENE-BIS-(5-(P-HYDROXY	JPMS-A-0055-0028
HEXANEDIONE) AND 2,2-	ALDOSYLIDENE-BIS-(5,5-DI METHYL-1,	JPMS-A-0055-0028
PHOTOMETRIC ANALYSIS OF	ALDRIN AND DIELDRIN IN A TECHNICAL	CANH-A-0010-1253
RING THE MANUFACTURE OF	ALDRIN.= + COMPOUNDS OBTAINED DU	MGKF-A-0072-0044
NTHOSINE FROM TRIFOLIUM	ALEXANDRINUM SEEDS.= XA	CUSC-A-0035-0011

Figure 18
Sample of KWIC Index to Chemical Titles



CHEMICAL TITLES

BIBLIOGRAPHY

J. Clin. Invest., 45, No. 2 (1966)

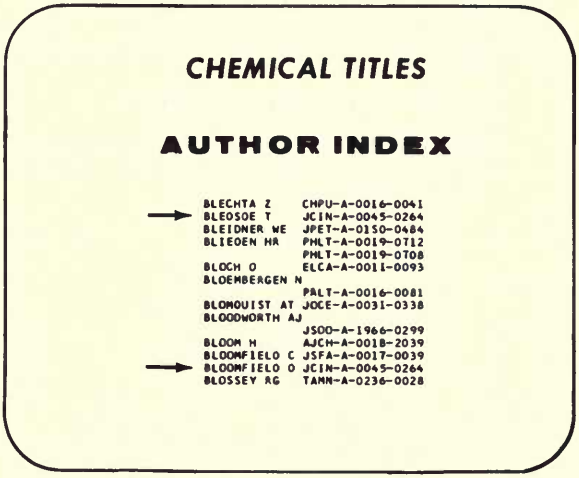
JCIN-A-0045-0228 PORTE O GRABER AL AUZUYA T
WILLIAMS AH
EFFECT OF EPINEPHRINE ON IMMUNO REACTIVE INSULIN LEVELS IN MAN.* 228-38

JCIN-A-0045-0249 GROSS AE SCHRCEDER EA GABRIO BW
PYRIDINE NUCLEOTIDES IN ERYTHROCYTE METABOLISM.* 249-55

→ JCIN-A-0045-0264 BLEOSOE T LIDDLE GW RIONDEL A
ISLAND DP BLOOMFIELD O SINCLAIR-SMITH P
COMPARATIVE RATES OF INTRAVENOUSLY AND ORALLY ADMINISTERED ALDO STERONE.* EVIDENCE FOR EXTRA HEPATIC FORMATION OF ACID HYDROLYZABLE CONJUGATE IN MAN. 264-9

JCIN-A-0045-0270 AEGAN TJ BORDAEMIDIS G MOSCHOS ER
OLDEWURTEL HA LEHAN PH HELLEMS HK
ACUTE METABOLIC AND HEMODYNAMIC RESPONSES OF THE LEFT VENTRICLE TO ETHANOL.* 270-80

Figure 19
Sample of Bibliography Section of Chemical Titles



CHEMICAL TITLES

AUTHOR INDEX

→ BLECHTA Z CHPU-A-0016-0041
→ BLEOSOE T JCIN-A-0045-0264
BLEIDNER WE JPET-A-0150-0484
BLIEDEN HA PHLT-A-0019-0712
PHLT-A-0019-0708
BLOCH O ELCA-A-0011-0093
BLOEMBERGEN N
PALT-A-0016-0081
BLOMQUIST AT JOCE-A-0031-0338
BLOODWORTH AJ
JSOO-A-1966-0299
BLOOM H AJCH-A-0018-2039
BLOOMFIELD C JSFA-A-0017-0039
→ BLOOMFIELD O JCIN-A-0045-0264
BLOSSEY AG TAMM-A-0236-0028

Figure 20
Sample of Author Index to Chemical Titles

CHEMICAL-BIOLOGICAL ACTIVITIES

ILLUSTRATIVE DIGEST

J. Biol. Chem., 241, No. 2 (1966)

259 STUDIES ON THE TRIPHOSPHOPYRIDINE NUCLEOTIDE-CYTOCHROME f REDUCTASE OF CHLOROPLASTS.
JBCH-A,241,2,66,279-85
Zanetti G, Forti G;Univ. Milan.

The kinetic properties of a spinach LEAF CHLOROPLAST crystalline flavoprotein enzyme, both as TPNH-cytochrome f reductase and as TPNH diaphorase, as well as its molecular weight, content of
146145 FAD, and reductive inactivation, were studied.

259-1 Kinetic reaction studies of the activity of TPNH-cytochrome f reductase gave a Km for the coenzyme of 2.1 micromolar, suggesting the formation of a ternary complex of
53576 TPNH,
Cytochrome f, and the enzyme.

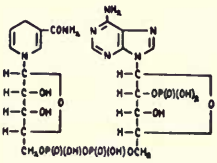
259-2 The REDUCTION of ferricyanide by the enzyme gave a Km of 30.7 micromolar and involved the reaction of TPNH with the enzyme to form
53558 TPN and reduced enzyme followed by the reaction of the reduced enzyme to form ferrocyanide and oxidized enzyme.

259-3 The DIAPHORASE ACTIVITY of the enzyme was reversibly inactivated by preincubation with TPNH; TPN prevented and reversed this INACTIVATION.

259-4 57136 Urea irreversibly inactivated DIAPHORASE ACTIVITY of the enzyme and converted the reversible inactivation by TPNH into an irreversible process.

259-5 The activity of TPNH-cytochrome f reductase was not effected.

259-6 The functioning of FAD and the maintenance of bonds sensitive to Urea are essential to the preservation of the native structure of the enzyme required for activity.



53576

Figure 21
Sample Page of the Digest Section of CBAC

CHEMICAL-BIOLOGICAL ACTIVITIES

KEYWORD-IN-CONTEXT INDEX

253-7	OCHONDRIA fractions+ When the	CHLOROPLAST and MITOCHONDRIA frm
→ 259	properties of e spinech LEAF	CHLOROPLAST crystelline Fievopro
387-3	ORYLATION decreased to zero in	CHLOROPLAST fragments from LEAVE
36	.=	CHLOROPLAST STRUCTURE AND FUNCTI
387-6	STRUCTURAL INTEGRITY of the	CHLOROPLAST, permitting the fine
8	ADP PHOTOREDUCTION by isolated	CHLOROPLASTS by 200-350%.= + p
286-1	factor was removed from the	CHLOROPLASTS by washing or hypot
383	bonate, isolated whole spinach	CHLOROPLASTS formed a small amou
36-10	distinguishable from those+	CHLOROPLASTS from CALLUS past th
431	REASED PHOTOPHOSPHORYLATION BY	CHLOROPLASTS FROM CHINESE CABBAG
431	ee cabbage, in comparison with	CHLOROPLASTS from control LEAVES
431-4	Y MECHANISM that would prevent	CHLOROPLASTS from functioning ei
36-9	d in comparison with+ The	CHLOROPLASTS from growing CALLUS
387-1	Pyridine nucleotides in the	CHLOROPLASTS from immediately de
387-1	at 80 and 20%, respectively,+	CHLOROPLASTS from LEAVES floated

Figure 22
Sample of KWIC Index to CBAC

CHEMICAL-BIOLOGICAL ACTIVITIES

MOLECULAR FORMULA INDEX

DIGEST NUMBER	MOLECULAR FORMULA	REGISTRY NUMBER
5	$C_{21}H_{30}N_7O_{17}P_3$	53576
20-4		
22-5		
77-11		
→ 259-1		
297-3		
411		
82	$C_{21}H_{30}O_2$	57830
166-1		

Figure 23
Sample of Molecular Formula Index to CBAC

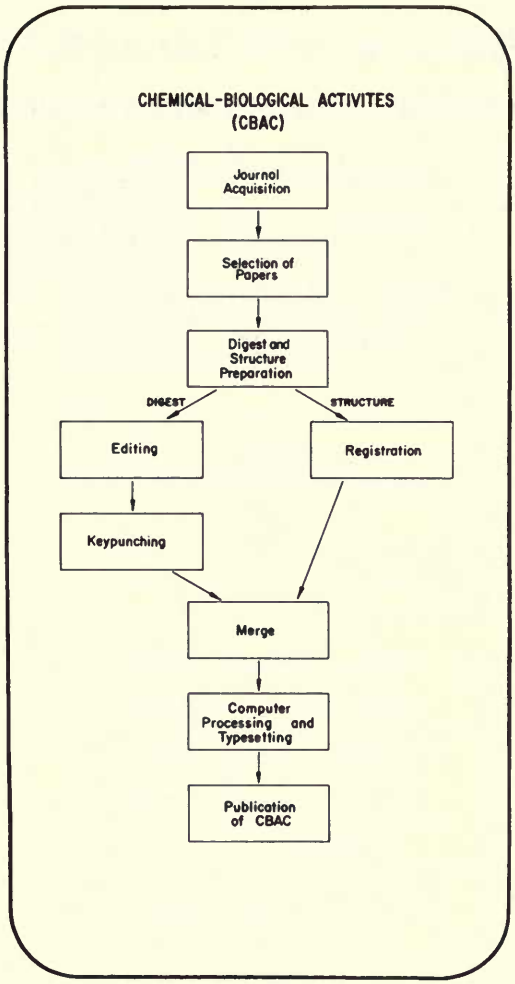


Figure 24
Flow Chart for the Preparation of CBAC

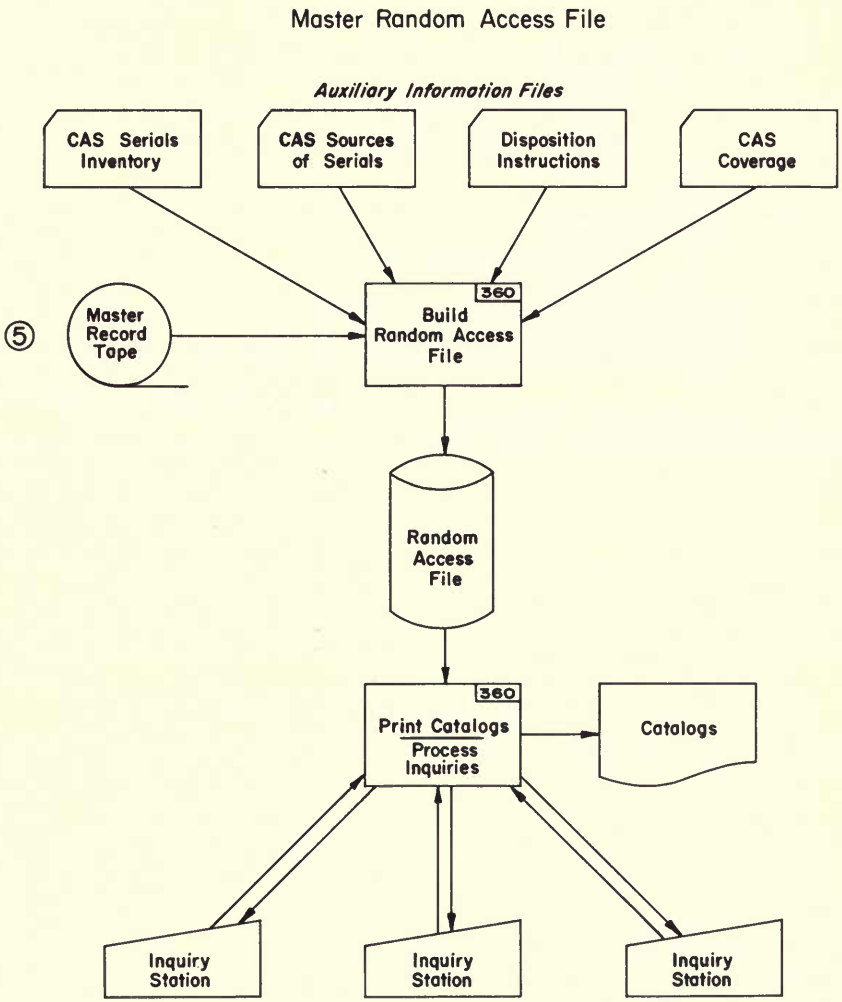


Figure 25
Flow Chart for Input and Output of Master Random Access File

Appendix 2

CAS Computer Configurations

1. IBM 360 Model 40
   - 32 K storage
   - 4 7-channel tape drives (30 KB tapes)
   - 1 9-channel tape drive
   - 2 printers
2. IBM 7010
   - 80 K storage
   - 8 7-channel tape drives
   - 1301 disk file
3. Two remote terminals—IBM 1050's
4. Increase in storage capability of 360 Model 40 to 128 K—March 15, 1966
5. Full operating system for the 360 Model 40, disk file and floating point arithmetic—April 1966