Effects of Mechanization on a Chemical Information Service*

A. W. ELIAS, G. S. REVESZ, and G. H. FOEMAN†
Institute for Scientific Information, 325 Chestnut St., Philadelphia, Pa. 19106

Received January 23, 1968

The *Index Chemicus*, an abstracting and indexing service devoted to reporting new chemical compounds and their synthesis (a registry of all new chemical compounds), evolved to its present form through a series of incremental mechanization steps. These procedures produce effects on indexing, editing, printing, format, and timing of the publication. The paper recounts the chronologically arranged mechanization plan, its theoretic and systematic basis, indicating the specific benefits, problems, by-products, and future plans.

The development of a chemical information publication or service is dependent on a number of diverse factors and pressures (Figure 1), such as economic pressures for sales or other acceptance, financial factors for control of production costs, alteration and addition of new information elements, expansion of coverage, and technical needs for revision. There are also timing demands which paradoxically require more features and coverage in *less* time.

When these pressures have to be met under economic constraints and realistic information manpower restrictions, only planning and care in the design of the system can hope to achieve near optimal results. It should also be realized that 100% effectiveness in such planning is improbable. Thus, a system or plan which provides "reasonable" exception procedures probably will be most successful in an informational and commercial sense. By exception, we mean the design of a system, plan, or procedure that copes effectively with those events met most often, recognizing the exceptions, and shunting them to a special handling status where they are specifically dealt with as encountered.

The *Index Chemicus* (IC) service of the Institute for Scientific Information, now entering its ninth year of publication, has grown with active adaptation to the needs and pressures of the chemical community and the economics of commercial publishing. It has done so by maintaining a flexible and responsive system design coupled with a continuing program of evaluation. These allow for changing information needs of subscribers and provide for the introduction of improved mechanization techniques.

An information service begins with the recognition of the basic needs of the users it is to serve. In the case of the original *Index Chemicus* in 1960–61, the primary objective, in response to these needs, was timely reporting on the synthesis of new compounds

By concentrating on speedy reporting, IC could not examine all the thousands of journals in the chemical literature, and so the first exception procedure was developed. It was decided that the reports of *new* compounds constituted an important and specially useful subset of the total chemical literature. By analysis of the literature it was determined that, of the thousands of journals, a restricted number reported on the majority of new chemical compounds. While the number of these journals has increased over the nine years of IC, this basic coverage philosophy remains unaltered.

The second need was to determine novelty. It was difficult to find a definition of a "new" compound that was acceptable to all. Since the primary audience for the new publication was the research chemist, it was determined that the normal criteria employed by the primary publication was best. Thus the *Index Chemicus* listed, as it does today, compounds for which the author does not cite a prior reference. Conversely, a lack of novelty is established by the author's reference to a previous synthesis of the same compound.

Before discussing the achievement of prompt coverage, it is necessary to develop the extent of the information to be reported, since indexing and preparation of an abstract depends on its content. The task of *Index Chemicus* was two-fold. It was, and is, a current awareness publication as well as a retrospective searching tool. The user needs to scan current work. To satisfy the retrospective searching need, cumulations on a quarterly and annual basis were contemplated. A basic systems decision was made which has profoundly affected subsequent developments. This was to develop the index entries from elements present in the abstract itself. The policy provides for two-fold use of the same inputs and has been implemented all along the line as will be seen later.

The concept of the *Index Chemicus* abstract was developed with great care, and after thorough scrutiny of the primary source—the journal articles—six major elements were established as being necessary. The first of these is the *bibliographic* description: title and citations. The second is *personal* information: the authors and their locations. In chemical articles there is, of course, a *legal* aspect to be considered and so the date that the manuscript is received by the journal for purposes of effective disclosure and the novelty status are the third element. Fourth, there is the *compound* oriented information. This includes such data elements as the structural diagram,

*Presented in Symposium on "Automation of Information Operations," ACS Middle Atlantic Regional Meeting, Philadelphia, Pa., February 1, 1968

†Present address, Information Interscience Incorporated, Philadelphia, Pa.

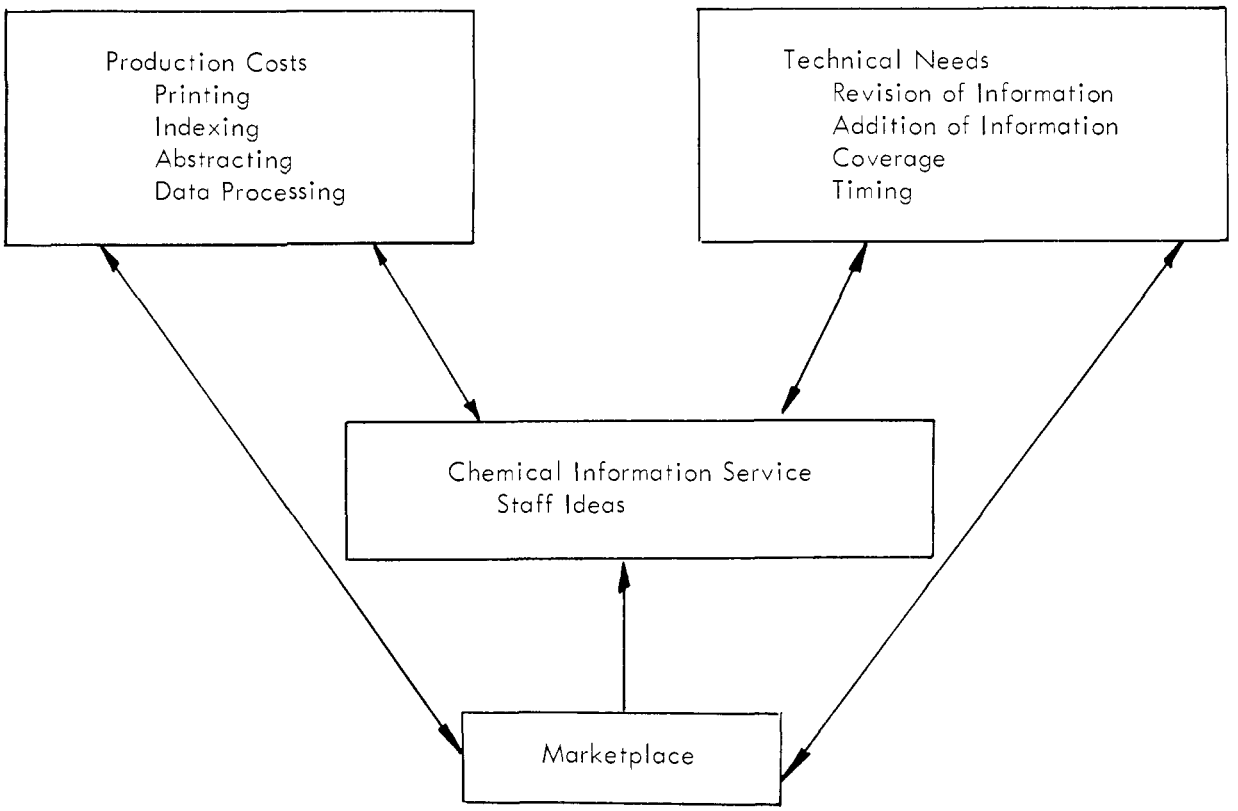


Figure 1. Factors affecting chemical information services

the molecular formula, and the names of the compounds prepared. Fifth, there is the *registry* or lookup-information needed to identify uniquely the published abstract and compound with the journal article. The sixth segment is the text itself and any version of it, such as author abstract or summary, which may stand by itself as representing the content of the article.

From this analysis it was determined to base the *Index Chemicus* abstract on the following:

Bibliographic—Article title in English; Journal Name, Volume, Pagination, and Year; Language of original article
Personal—Author name(s); Institution where work was performed (address) and mailing address of first author if different from above
Legal—Implied novelty statement; Receipt date of manuscript by journal
Chemical—Molecular formulas; Structural diagrams; Chemical names
Registry—Unique serial number for each abstract, and within each abstract a serial number for each compound; The author's diagram, reference number, and/or name for each compound; Page number for each compound

From these abstract elements we selected the molecular formulas, the author(s), and the journals processed as the *index* entries. It will be noted that nothing of the text was used, nor was any attempt made to provide a subject or topical index entry. As a compound oriented publication, the main lookup was based on the molecular formula itself, with only the author(s) name as a back-up.

The timing requirements for abstracting and indexing are factors in establishing the frequency of the publication. With 114 journals to be covered in 1960–61, we could estimate the number of issues, articles, and pages to be examined, which in turn determined that the work force could do the necessary abstracting in about 30 working days. However, this was not the only basis for deciding on a monthly publication. In addition to considering the printing cycle, it was necessary to decide on how much information could be assimilated in the time available to a working chemist.

Having established the purpose, philosophy, and content of the product and its indexes we next turned to the production system and the use of machines in that system. Figure 2 shows the 1960–61 *Index Chemicus* system flow chart and the changes made in it over the years. The journals go through a normal accession procedure where their arrival is recorded, necessary claims are made, and the issues routed to the product offices. A scanning operation is performed based on the novelty requirement for abstracting. Those journals which have no articles containing new compounds are processed no further except for the fact that the notation "NNC" (No New Compounds) is made for the journal index which is prepared by typing. This is done at the point where it is known exactly which journals are going to be included in a given IC issue. Those journals containing new compounds enter the abstracting phase. While shown in two steps, the scanning and abstracting processes are contiguous. The same person scanning the issue performs the abstracting task. At this point the compounds to be indexed are noted, their molecular formulas checked, and any needed translation of titles is made. The page numbers

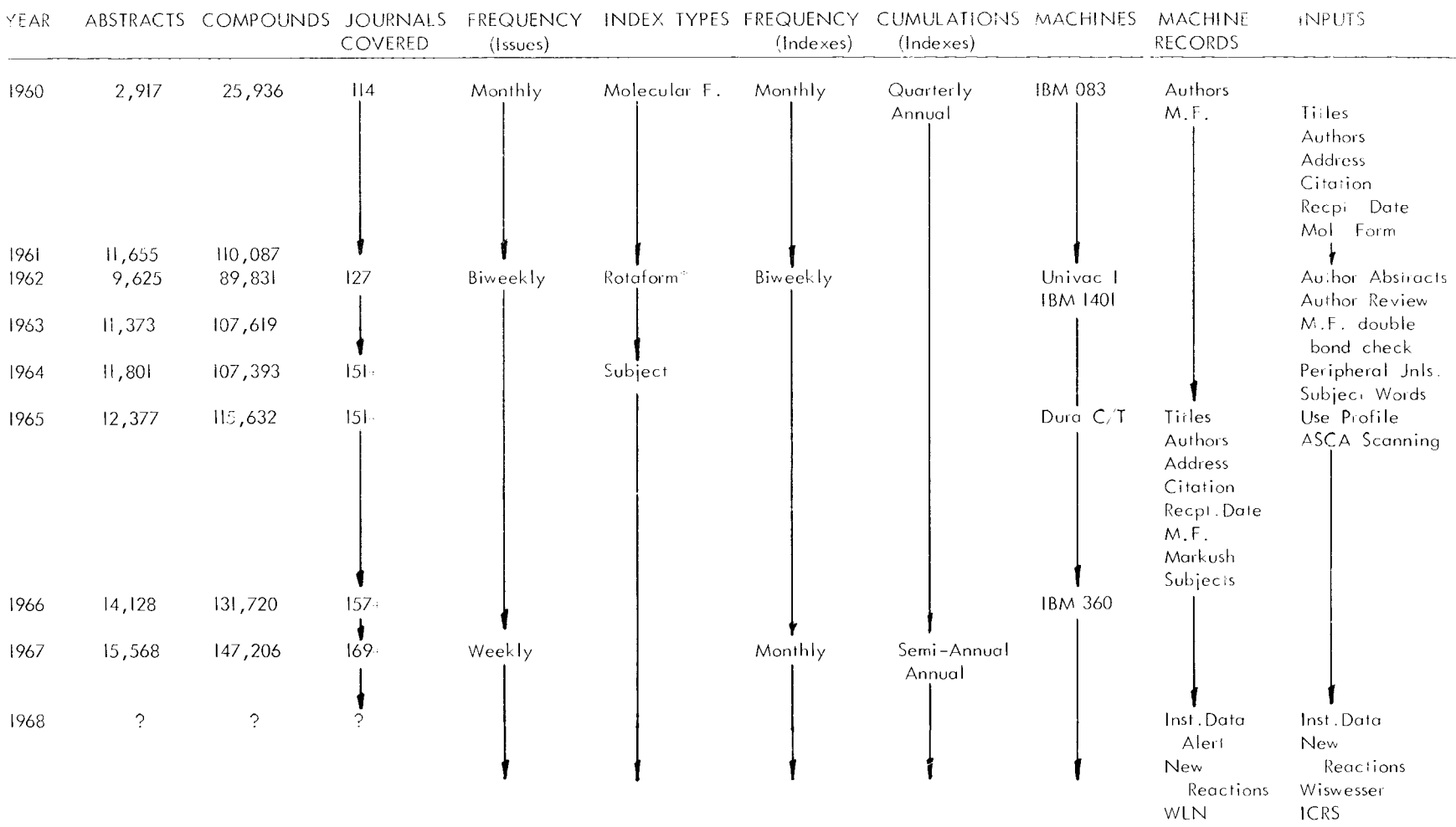


Figure 2. Production changes

and diagram references are indicated for each compound and a notation given of which diagrams are to be used. If additional molecular formulas or diagrams *not* provided by the author are needed, they are roughly drawn. At this point, a preliminary edit of the abstract is performed and the material is sent to a typist who prepares the abstract text—that is the title, authors, citation, address, receipt date, and molecular formulas. Diagram reference numbers are also typed as well as any compound names. This record and the original journal is then given to an artist who prepares the necessary diagrams and arranges the information in IC format. A copy of the abstract is now sent to the article's author for his approval. The IC pages with returned author proofs attached are subsequently sent to the abstracting group for proofreading to avoid inadvertent inaccuracies caused by the location of information in a poorly juxtaposed manner, typographical errors, or incorrect diagrams. Any alterations due to the proofreading step are made in the final art phase. It is important to note that the unique numbers for each abstract have not yet been applied. Since these are sequential, they cannot be affixed until the content of the issue has finally been established. This comes later and is influenced by printing economies to be gained by running the optimum number of pages. Any corrections noted on the author's proof too late to be used in the issue are used for correcting the index cumulations. At this point, the issue begins to take shape. The number of pages is established and the number of index pages estimated. The numbered copy is keypunched, verified, and sorted with an 083 sorter for the molecular formula and author indexes. These are printed by a 407 tabulator, pasted up in a six-column format by the art group, and sent to the printer.

It is now interesting to note what effect the use of data processing machinery has had. Observe that very little information is really being recorded in machine form. All of the processing steps are done by hand, including the original entry of the molecular formula and author data. They are captured in machine form *afterwards*, and the machines are employed only to alphabetize and print the index. Thus, certain information is being keyed twice and the machines are used in a minor and peripheral way. Figure 2 shows the changes that took place in 1962. It will be noted a computer, the Univac 1, has replaced the tabulating equipment, most particularly the 083. Computer sorting is more efficient and consequently cheaper as the number of items to be sorted increases. As the biweekly indexes increased in size, so did the quarterly and annual cumulations. In addition, by programming the Univac, the need for fixed field punching was eliminated. Both the encoding of the formulas and their punching became easier, and certain editing could be accomplished on the computer. Still further improvements were made in the index formats possible on the computer printer and not possible on the 407.

It was in 1962 that the first text element was added by picking up available author abstracts *in toto* and adding them to the abstract information. No machine record was made of this information. In 1962 also, the results of the author proof operation were added to the abstracts to indicate the validation of the material by the author. Because of this procedure, the data in *Index Chemicus* are sometimes more accurate than the original journal article.

In mid-1962, another computer change was made. The IBM 1401 replaced the Univac basically for the reason that the 1401 was more readily available and also because the quality of the printouts from the 1401 were superior to the Univac 1. At this time, *Index Chemicus* went biweekly in response to the increase in the number of items to be indexed.

At the end of 1962, the first return on the investment in machine language recording was made. By using the molecular formula information and subjecting it to a program that rotated the entries to provide the Rotaform molecular formula index, retrieval was facilitated since the user could enter the index on any element in a compound. No additional input cost was incurred so that only computer time, pasteup, and printing were needed to produce the Rotaform (*1, 2*).

In 1963, the quality of the input was increased by subjecting all molecular formulas to routine double bond calculation. The *Index Chemicus* was again put in the position of being more accurate than the original journal article.

In mid-1964, a production problem began to intrude. As you will recall, all punching took place after the decision was made as to which material was to be included in IC. This meant that deadlines for the punching were very tight, as were allotted times for corrections. Therefore, a simple change was made in the numbering procedure. As shown on Figure 2 use was now made of arbitrary numbers. These were attached to the abstracts as they were produced. The abstracts were sent for keypunching immediately, eliminating the pileup of material at the very end of the process. As it was decided which material was to go in a given issue, the arbitrary numbers were removed, replaced by sequence numbers, and the cards renumbered.

In 1964, the first major change in indexing entries took place. As a result of user surveys and editorial board recommendations, a subject index was added. The words used to characterize each abstract were integrated with the abstract information and then processed by computer to form the index. The integration of words and abstracts was achieved by having the words lead to abstract numbers rather than IC page numbers.

At this time, a further source of new compound input was instituted. Although the original premise held of most new compound publications occurring in a restricted number of journals, the *Index Chemicus* staff now routinely scanned the titles of all contents pages appearing in *Current Contents*, determined those which possibly reported new compounds, and subjected these titles to indexing as appropriate. In this way the coverage now reflected an additional 1300 journal titles.

In 1965 another major change was effected in the mechanization of the production of *Index Chemicus*. This was done in response to two main considerations. First, economics: the double keying operation, where the abstracts were first typed and then certain elements were keypunched involved duplication of effort of selected parts of the abstract. This, plus the fact that this mode of production left no machine record for the rest of the information, stimulated a review of the production

INVESTIGATION IN THE 4-PHENYLPIPERIDINE SERIES. VII. DERIVATIVES OF N-SUBSTITUTED 4-PHENYL-4-FORMYLPIPERIDINE AND 4-PHENYL-4-CYANOPIPERIDINE WITH ACTIVITY ON THE CENTRAL NERVOUS SYSTEM.
S. Chiavarelli, P. Mazzeo, A. Pugliese, M.A. Jorio, A.M. Russo.
Ist. Sup. Sanita, Rome, Italy. Recd. Apr. 30, 1964.
Farmaco Ed. Sci. 19(10),849-64(1964). In Italian

			X=	
1)	C25 H30 N2 O2	(A)	CH2	
2)	C26 H32 N2 O2	(A)	(CH2)2	(& di-HBr)
3)	C30 H40 N2 O2	(A)	(CH2)6	(& di-picrate)
4)	C34 H48 N2 O2	(A)	(CH2)10	(& di-HCl)
5)	C32 H36 N2 O2	(A)	CH2-C6H4-CH2	(& di-HCl)
6)	C21 H23 N O2	(III)	(as HCl)	
7)	C21 H22 CL N O2	(IV)	(as HCl)	
8)	C21 H22 Br N O2	(V)	(as HCl)	
9)	C21 H22 F N O2	(VI)	(as HCl)	
10)	C22 H25 N O2	(VII)	(as HCl)	
11)	C22 H25 N O3	(VIII)	(as HCl)	
12)	C23 H27 N O3	(IX)	(as HCl)	
13)	C21 H22 N2 O	(X)	(as HCl)	
14)	C21 H21 CL N2 O	(XI)	(as HCl)	
15)	C21 H21 Br N2 O	(XII)	(as HCl)	
16)	C21 H21 F N2 O	(XIII)	(as HCl)	
17)	C22 H24 N2 O	(XIV)	(as HCl)	
18)	C22 H24 N2 O2	(XV)	(as HCl)	
19)	C23 H26 N2 O2	(XVI)	(as HCl)	
20)	C25 H28 N4	(XVII)		
21)	C19 H20 N2 O2 S	(B)		
22)	C19 H21 N O3 S	(C)		

CNS ACTIVITY

Figure 3. Dura card to tape printout

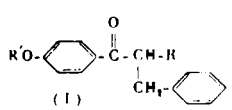
47751

1-(4-HYDROXYPHENYL)-3-PHENYLPROPANOL DERIVATIVES WITH A BASIC GROUP IN THE 2-POSITION.
G. Bramanti, G. Dell'Omodarme, G.F. DiPaco.
Lab. Guidotti & C., Direzione Ric., Pisa, Italy. Recd. May 15, 1964.
Farmaco Ed. Sci. 19(10),837-48(1964). In Italian

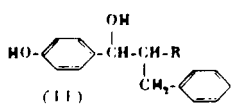
A description is given of some new basic derivatives of 1-(4-hydroxyphenyl)-3-phenylpropanol bearing an –NRR' group in position 2 where R and R' are hydrogen, alkyl, aralkyl or are integrated into a heterocyclic nucleus.

All these amines were characterised both as bases and as hydrochlorides. Depending on the compound in question, one of two synthetic routes was used to prepare the related ketone which was then catalytically reduced to the desired product.

The compounds were tested pharmacologically for activities suggested by their structure and the results of this preliminary screening are reported.



(I)



(II)

				R'=	
1)	C15 H15 N O2	(I)	NH2	H	(& HCl)
2)	C16 H17 N O2	(I)	NH-Me	H	(& HCl)
3)	C17 H19 N O2	(I)	NH-Et	H	(& HCl)
4)	C18 H21 N O2	(I)	NH-Pr-i	H	(& HCl)
5)	C19 H23 N O2	(I)	NH-Bu	H	(& HCl)
6)	C22 H21 N O2	(I)	NH-benzyl	H	(& HCl)
7)	C17 H19 N O2	(I)	N(Me)2	H	(& HCl)
8)	C19 H21 N O3	(I)	morpholino	H	(& HCl)
9)	C20 H23 N O2	(I)	piperidino	H	(& HCl)
10)	C19 H21 N O2	(I)	pyrrolidino	H	(& HCl)
11)	C22 H18 O3	(I)	H	Bz	
12)	C22 H17 Br O3	(I)	Br	Bz	
13)	C15 H13 N O3	(I)	NO	H	
14)	C15 H17 N O2	(II)	NH2		(& HCl)
15)	C16 H19 N O2	(II)	NH-Me		(& HCl)
16)	C17 H21 N O2	(II)	NH-Et		(& HCl)
17)	C18 H23 N O2	(II)	NH-Pr-I		(& HCl)
18)	C19 H25 N O2	(II)	NH-Bu		(& HCl)
19)	C22 H23 N O2	(II)	NH-benzyl		(& HCl)
20)	C17 H21 N O2	(II)	N(Me)2		(& HCl)
21)	C19 H23 N O3	(II)	morpholino		(& HCl)
22)	C20 H25 N O2	(II)	piperidino		(as HCl)
23)	C19 H23 N O2	(II)	pyrrolidino		(as HCl)

USE PROFILE

ANTIBRONCHOSPASMODIC AGENTS
RESPIRATORY STIMULANTS
VASODILATORS

Figure 4. *Index Chemicus* abstract (1965) showing use profiles

methods. Typing had been employed for esthetic reasons—getting sharp fonts and enabling the use of upper and lower case characters. This was thought especially important with regard to some chemical elements in molecular formulas such as "I." When all uppercase was used close to numbers, it was felt possible confusion might result such as might occur with a "I" and a "2" yielding an apparent "12." The need for upper and lower case printing was thought important, but the upper/lower case fonts on computers were less than satisfactory. Still, if machine recording of the information with formatting left to the machines was possible, keypunchers could replace skilled typists. A solution was found through the use of a Dura typewriter and a card to paper tape converter as shown in Figure 2. By placing *all* the information except the diagrams in machine language, it was available for reprocessing and the entire input of IC could be stored more efficiently and possibly sold as a magnetic tape product. At this point, there were two ways to go. We could use punched paper tape, convert it to cards, and then run the cards through the computer. Or we could use cards, convert them to tape to drive the typewriter, and then run them in the computer for the indexes. We selected the card to tape method because error correction seemed easier in the unit card record, and because the cards could be produced on existing equipment and run off-line (after conversion to tape) on a Dura typewriter. The formatting took place in the card conversion step, with minimal fixed field requirements once the card type information had been fed to the machine. As a result of this modification, copy was generated after one keystroking operation and all information was captured in machine language. The new subject index cards were integrated into the system so that they printed out at the bottom of the abstracts (Figure 3). This revealed the assigned subject terms to the IC staff editors and enabled them to correct possible keying errors.

In 1965, another input element was added which may be thought of as a specialized form of index term. It was considered helpful to give the users a statement concerning the *uses* of the compounds. This use information was treated in two ways. When the article reported test data concerning the use of new compounds (whether positive or negative), the information was printed in the abstract itself as the "USE Profile" (Figure 4). The indication that the data are present was contained in a code which preceded the entry on the subject word assignment. When the use was a "potential" one, unsupported by any data, the words were assigned as subject index entries with the appended term "Potential." The user was then

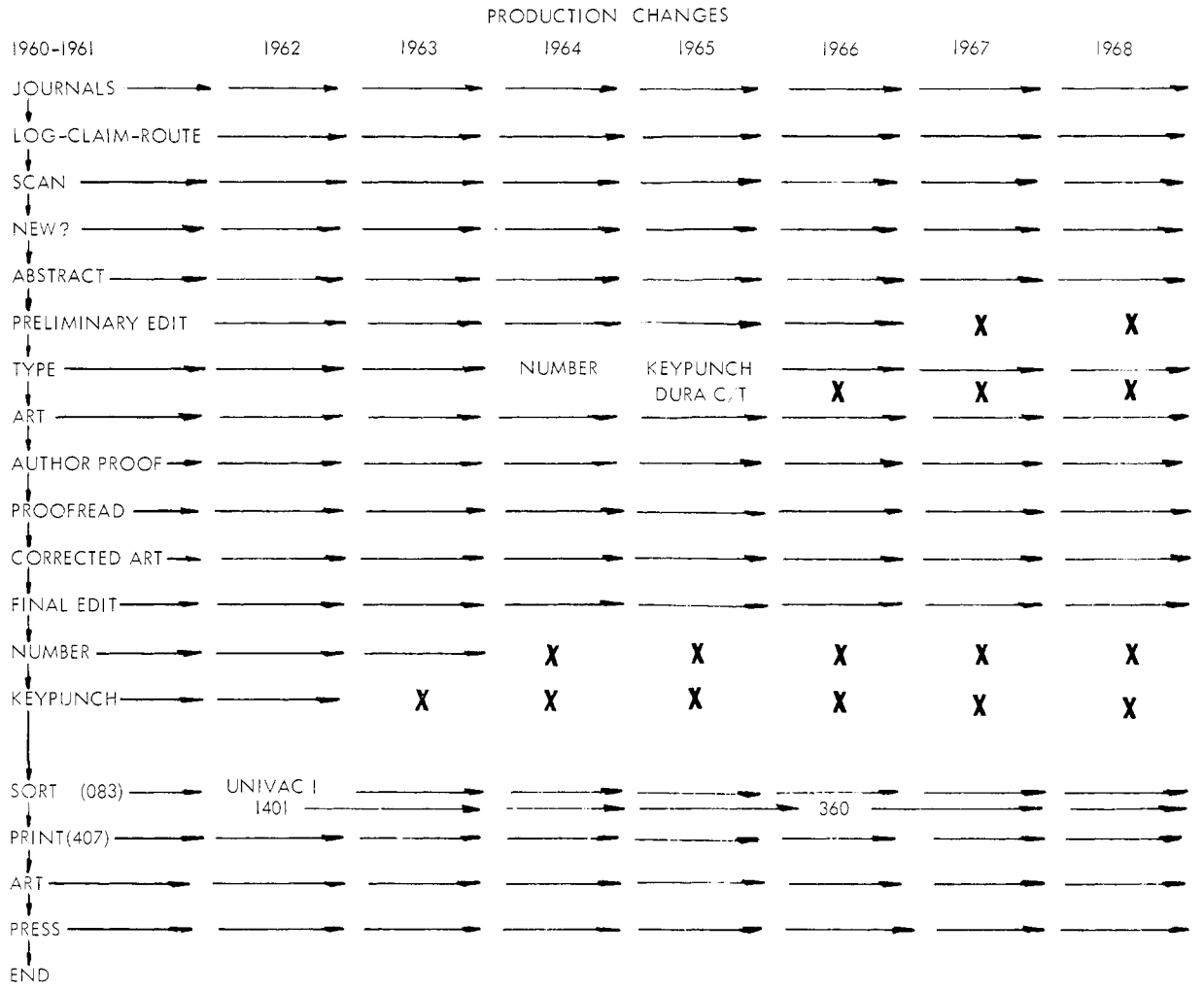


Figure 5. Chronological development of *Index Chemicus*

able to enter the subject index on any form of biological or industrial activity, find those articles reporting such compounds, and even find which have test data and which do not.

With the development of the ASCA system for selectively disseminating information, *Index Chemicus* became a major user of the service. Profiles were designed and ASCA scanned the titles, citations, etc. of all 1700 journals in the *Science Citation Index* for possible articles disclosing new compound information.

While the Dura procedure was effective and achieved the desired results, problems with the reliability of the card to tape converter led us to re-examine the basic premise for its use. This was, as will be recalled, primarily the need for upper and lower case characters. As an experiment, we tried producing an entire issue in upper case characters using the computer to format and print the abstracts. After a survey of our users and our own analysis of the appearance of all upper case, we eliminated the Dura apparatus and went to the computer. Now, the cards simultaneously produce the abstracts and the tapes for the indexes. At the same time as the upper case acceptability was tested, the Rotaform index was reevaluated. The response indicated that this index was receiving little use and so it was discontinued in 1965.

In 1966, the major change in IC production was the change to the IBM System 360 computer. There is little directly observable effect on the publication from the use of this machine; however, it provided an important ingredient in the next radical change which was effected in 1967. The year's basic coverage had been increasing as important new journals were appearing which disclosed new compounds. In addition, those publications we had started with kept increasing in size, frequency, and most important in the number of compounds disclosed (Figure 5). By 1966, we were indexing over 130,000 compounds annually and the size of the biweekly publication was becoming too voluminous for easy scanning. With the increase in efficiency and speed that the 360 allowed and with slight modification in editing procedures we were able to initiate *weekly* publication in 1967. Index frequency was adjusted since retrospective searching would be ill served by having users scan 52 weekly issues if weekly indexing was provided. It was decided that monthly, semi-annual, and annual cumulations of indexes provided the best approach and this was used in 1967. The user must scan at most five monthly indexes until the semi-annual is delivered and the semi-annual plus five monthlies until the annual is available. In 1967, the annual cumulation was delivered before the end of calendar 1967.

Changes continue to be made. After further studies and user surveys, *Index Chemicus* turned its attention to additional information elements for chemistry. Two important additions have already been made in 1968. These are the Instrumental Data Alert and New Reaction Display. By scanning for the presence of different types of instrumental analysis data and any disclosure of new reaction schemes, *Index Chemicus* widens its utility still more.

Yet another opportunity to expand the *Index Chemicus* system is coming about in 1968. The new *Index Chemicus* Registry System is adding substructure searching capability to the file by employing Wiswesser Line Notations for every compound. This feature will be available only to subscribers of *Index Chemicus* in magnetic tape form. Several forms of WLN printouts will be supplied as will a complete tape version of the machine data from the *Index Chemicus* itself. Through the Wiswesser notation the detailed chemical structures now become available. Atom-by-atom searches are feasible, alone or in combination with any of the other informational ingredients available on the IC tapes (*3*).

Figure 5 shows the chronological development of *Index Chemicus*. Not a year has gone by without a change, adjustment, or revision. Coverage has increased 48% and frequency four-fold; compounds, articles, information, indexes—every element has been improved through the utilization of machines for those functions that machines do best, leaving people free for the intellectual tasks that only they can perform. These things are possible when the system is approached as an open-ended tool for dynamic and responsive service to chemical documentation.

LITERATURE CITED

(1) Garfield, E., "Generic Searching by Use of Rotated Formula Indexes," Division of Chemical Literature, 141st Meeting, ACS, Washington, D. C., March 1962; J. CHEM. DOC. **3**, 97 (1963).

(2) Sher, I., J. O'Connor, and E. Garfield, "Rotadex—A New Index for Generic Searching of Chemical Compounds," Division of Chemical Literature, 143rd Meeting, ACS, Cincinnati, Ohio, January 1963; J. CHEM. DOC. **4**, 49 (1964).

(3) Barnard, A. J., and W. J. Wiswesser, "Computer-Serviced Management of Chemical Structure Information," *Laboratory Management* **5** (10), 34–6, 38, 40–2, 44 (1967).

(4) Garfield, E., and G. Foeman, "Statistical Analyses of International Chemical Research by Individual Chemists, Languages, and Countries," Division of Chemical Literature, 148th Meeting, ACS, Chicago, Ill., September 1964.

(5) Elias, A., E. Garfield, G. Foeman, and G. S. Revesz, "Mechanization of Chemical Information Publications and Services," Division of Chemical Literature, 149th Meeting, ACS, Detroit, Mich., April 1965.

(6) Sher, I., A. Elias, and E. Garfield, "Control and Elimination of Error in ISI Services," Division of Chemical Literature, 151st Meeting, ACS, Pittsburg, Pa., April 1966; J. CHEM. DOC. **6**, 132 (1966).