Need for Primary Periodicals As Determined by SDI

BARBARA STEFANIAK

Institute for Scientific, Technical and Economic Information, Warsaw, Poland

Received March 31, 1980

An investigation of primary periodicals that appeared in references selected from *CACon* was carried out on the output for about 2000 profiles. Frequency lists, in alphabetical and rank order, were created for the data obtained in the course of routine processing of two volumes: 86 and 89 of *CACon*. In both issues, a relatively small number of journals, e.g., 212 and 220 out of 6209 and 7039 titles, respectively, supplied 50% of the relevant papers for SDI users. The method is described, and the results are compared with the *CASSI* list of journals most frequently cited in *Chemical Abstracts*.

INTRODUCTION

Implementation of efficient information services, which supply information selected from comprehensive bibliographic data bases, covering the world's scientific and technical literature, increases demands of users for primary sources. The experience of many libraries, including some in Poland where an SDI system was introduced to users in late 1974, has shown deficiency in collections, less evident before, as well as lack of good location tools, like complete catalogs or union lists. Data base searching has a direct impact on libraries—it may affect the acquisition policy of the library as well as the interlibrary loan traffic.[1] One of the major faults of which data base search systems are accused is that they identify articles but provide no way for the user to obtain them.[2,3] An interesting solution was accomplished by coupling the computer-based reference services and supporting document delivery system at the University of California.[4] Ready access to the copies of documents found in the output is considered to be extremely important for the user if an SDI service is to be effective; so first of all one must know the sources cited in an SDI output, then determine their presence or absence in the libraries accessible to the user, and finally, if needed, complete the collection.

There are many possibilities of data base use for purposes other than publishing and information retrieval, since much more than subject information is included in most of the data bases currently available. These data are useful for quantitative analysis of bibliographic features of scientific and technical literature.[5] There are some papers on unconventional use of data bases. *Science Abstracts* was used to make up a ranked list of journals containing papers on semiconductors by means of the DIALOG online information retrieval system.[6] The data base INSPEC was also used by the Acquisition Department of the Slovak Technical Library to search current references to papers presented at conferences, seminars, or meetings and to dissertations and other special documents. Retrieved information served as a useful tool to complete the collection.[7] Another study was carried out using five data bases—BIOSIS, CAIN, Index Veterinarius, MEDLARS, and PASCAL—at the School of Veterinary Science, University of Hanover.[8]

The subject of the present study is the demand of SDI service users for primary documents published in periodicals included in *Chemical Abstracts Condensates* and cited in the output.

The idea of this approach is as follows:

(1) **Search Profile.** This is a computer-readable description of a user field of interest, representing user's information needs in relation to the problems being worked upon.

(2) **SDI Output.** This consists of references retrieved from the data base according to the user's interests (profile), representing needs regarding primary source documents.

METHOD

There are different methods, employing ranking techniques, used as aids to journal selection in order to satisfy the demands of library users—correlation between results obtained by using them is low.[9–11] Some methods are based on user activity (user studies and user judgement)[12–15] or on citation analysis,[16–18] others on size or productivity of primary journals, evaluated by means of abstracting journals.[19–21] Size and productivity rankings are usually concerned with one or the other formulation of the Bradford distribution. Only a few examples used data bases for these purposes.

The method applied in this study is based on the results of information retrieval obtained in routine SDI procedures. A similar approach has also been suggested by other Polish authors.[22] Periodical searches of current issues of data bases produce a set of references which make a specific "file of users' demand". The submitted method is relatively simple and consists of the following steps:

On the basis of all retrieved references, classified as "papers published in journals", following the processing of the tapes, abbreviated titles of periodicals are listed by means of the additional program, and frequency of occurrence of individual titles is recorded and cumulated.

At the end of an experiment, two frequency lists of journals are printed—one in alphabetical order and the other in descending rank order—both containing the number of citations of each title that occurred in the output.

The data in the frequency lists are analyzed to recapitulate the results and draw conclusions.

SCOPE

Information Center. The SDI system is sponsored by three institutions—the Warsaw Technical University, the Institute of Industrial Chemistry, and the Polish Academy of Sciences. The system is operated on a national scale by the Chemical Information Center of the Main Library, Warsaw Technical University.

Data Base and Period. The investigation was carried out on the references retrieved, on the basis of all SDI users' profiles, from two volumes of *Chemical Abstracts Condensates* (*CACon*)—volume 86 (January–June 1977) and volume 89 (July–December 1978).

Users and Profiles. SDI service subscribers, both individual and corporate users, are mostly from universities or engineering schools and from research institutes sponsored by the Ministry of Chemical Industry. In June 1977 and December 1978, information was retrieved from *CACon* based on 1758 and 2256 profiles, respectively. According to the results of a user

0095-2338/81/1621-0039$01.25/0

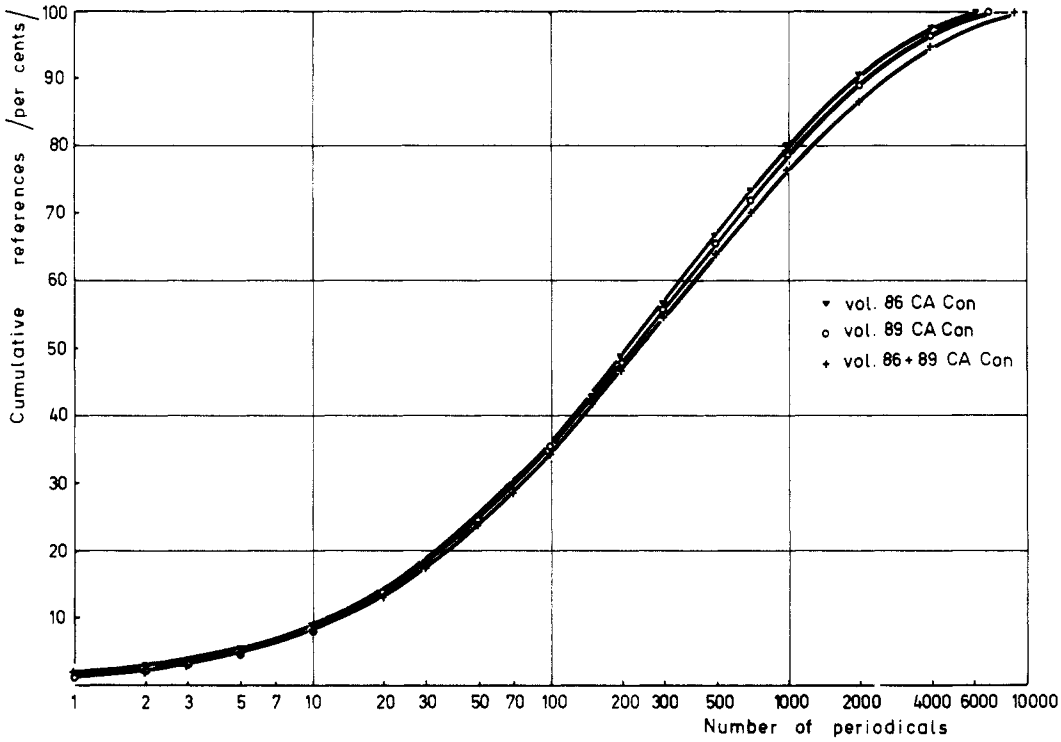

Figure 1. Distribution of selected papers among journals.

study in 1977,[23] four people, on average, make use of each delivered listing of references; so the number of individuals using SDI services might be estimated at 7000 to 9000.

RESULTS[24]

Number of Journals That Appeared in the Retrieved References. When the processing of volume 86 of *CACon* (26 tapes) was accomplished, frequency lists of abbreviated journal titles (cited in the references retrieved for the 1758 profiles) contained 6209 titles each, with *Journal of the American Chemical Society* in first place with 3060 citations. The list of journals (found in the listings for 2256 profiles) received from the tapes of volume 89 of *CACon* comprised 7039 titles, with *Analytical Chemistry* in first place with 2857 references.

When the alphabetical lists of journals cited in the references from *CACon* volumes 86 and 89 were compared and then merged, 4174 journals were found in both lists and 9074 titles in the united list. Thus more than half of the journal titles appeared only in the output from one of the two CACon volumes. That group consists of the journals in other domains, where chemistry is marginal to a main subject, as well as periodicals published infrequently or irregularly.

Number of References Retrieved and Their Distribution among Journals. According to the data of the Chemical Information Center, at the time of investigation (1st half of 1977 and 2nd half of 1978), 238 042 and 339 330 references were delivered to the users from volumes 86 and 89 of *CaCon*, respectively. References from periodicals amounted to 175 080 and 248 333 items, over 73% of the total in both cases.

The results of the analysis, presented in Table I, indicate that a comparatively small number of core journals that appear in the listings of references retrieved from the data base contribute a significant number of relevant papers for users.

Comparison of the lists of journals, arranged in rank order, obtained from two volumes of *CACon* indicates considerable overlap between titles (Table I, column 5). For instance, among journals that delivered 50% of papers, namely, 212 top titles selected from volume 86 and 220 top titles from volume 89 of *CACon*, there are 148 journals common to both, i.e., an overlap of about 70%.

Table I. Number and Percentage (in Parentheses) of Journals in Descending Rank Order (Columns 2–4), Yielding Specified Proportion (Column 1, Percent) of References Retrieved from Two Volumes of *CACon*, and Number of Journals (Column 5) Common to Ranked Lists Derived from Volumes 86 and 89

volume no. of *CACon* / no. of profiles / total no. of references	86 / 1758 / 175 080	89 / 2256 / 248 333	86 + 89 / 423 413	no. of titles common to both lists
10	12 (0.19)	13 (0.18)	14 (0.15)	5
20	36 (0.58)	35 (0.50)	37 (0.41)	22
30	72 (1.16)	70 (0.99)	76 (0.84)	45
40	128 (2.06)	128 (1.82)	138 (1.52)	83
50	212 (3.41)	220 (3.13)	234 (2.58)	148
60	346 (5.57)	372 (5.28)	397 (4.38)	252
70	577 (9.29)	636 (9.04)	685 (7.55)	446
80	1008 (16.2)	1073 (15.2)	1256 (13.8)	737
90	1995 (32.1)	2143 (30.4)	2577 (28.4)	1412
100	6209 (100)	7039 (100)	9074 (100)	4174

The nature of the numerical data comprised in Table I is a reminder of the regularities that occur in bibliography and were formulated by Bradford over 30 years ago as a "law of scatter" and then were developed as a Bradford distribution by using a simple graphical technique.[25]

The findings were transferred from numbers into graph form, and three diagrams (Figure 1) represent the relation between the cumulative number of references—papers selected according to users' needs—retrieved from volumes 86 and 89 of *CACon* (plotted on a decimal scale) and number of journals in descending rank order (plotted on a logarithmic scale).

The curves drafted on the basis of the results for the two volumes and the merged data prove to be of the same nature. Each can be divided into three zones. Initially the curves rise exponentially (the section of core journals) and then starting from about 65–70 journals become linear (cumulative number of references is directly proportional to the logarithm of journal number); finally, beginning at approximately 1100–1200, a droop from linearity is observed. The last section comprises a large number of journals that supply a relatively small

Table II. Journals That Provided SDI Users with 20% of Papers Relevant to Their Fields of Interest

rank of journal, N_{86+89} (N_{86}, N_{89})	journal title abbreviation	no. of papers each contributed n_{86+89} ($n_{86}+n_{89}$)	cumulative no. of papers $\Sigma n_{86,89}$
1 (1, 5)	*J. Am. Chem. Soc.*	4980 (3060 + 1920)	4 980
2 (3, 2)	*J. Chem. Phys.*	3988 (1714 + 2274)	8 968
3 (12, 1)	*Anal. Chem.*	3872 (1015 + 2857)	12 840
4 (4, 9)	*J. Chromatogr.*	3387 (1641 + 1746)	16 227
5 (2, 15)	*J. Organomet. Chem.*	3286 (1823 + 1463)	19 513
6 (6, 4)	*Kinet. Katal.*	3281 (1343 + 1938)	22 794
7 (26, 3)	*Bull. Chem. Soc. Jpn.*	2932 (688 + 2244)	25 726
8 (9, 18)	*Inorg. Chem.*	2491 (1108 + 1383)	28 217
9 (24, 7)	*Anal. Chim. Acta*	2479 (708 + 1771)	30 696
10 (7, 22)	*J. Org. Chem.*	2458 (1254 + 1204)	33 154
11 (10, 19)	*Zh. Prikl. Khim. (Leningrad)*	2438 (1107 + 1331)	35 592
12 (18, 11)	*J. Chem. Soc., Chem. Commun.*	2421 (784 + 1637)	38 013
13 (17, 12)	*Tetrahedron Lett.*	2408 (791 + 1617)	40 421
14 (5, 42)	*Chem. Phys. Lett.*	2401 (1561 + 840)	42 822
15 (34, 8)	*Zh. Anal. Khim.*	2345 (590 + 1755)	45 167
16 (22, 13)	*Surf. Sci.*	2324 (749 + 1575)	47 491
17 (49, 6)	*Izv. Akad. Nauk SSSR, Ser. Khim. Nauk*	2295 (508 + 1787)	49 786
18 (11, 20)	*J. Catal.*	2292 (1046 + 1246)	52 078
19 (16, 17)	*Elektrokhimiya*	2261 (820 + 1441)	54 339
20 (25, 16)	*Inorg. Chim. Acta*	2169 (706 + 1463)	56 508
21 (8, 36)	*Dokl. Akad. Nauk SSSR*	2124 (1235 + 889)	58 632
22 (80, 10)	*Can. J. Chem.*	2101 (359 + 1742)	60 733
23 (19, 21)	*Zh. Fiz. Khim.*	2000 (781 + 1219)	62 733
24 (13, 27)	*J. Phys. Chem.*	1953 (909 + 1044)	64 686
25 (23. 23)	*Fresenius' Z. Anal. Chem.*	1918 (740 + 1178)	66 604
26 (14, 28)	*Zh. Obshch. Khim.*	1898 (881 + 1017)	68 502
27 (107, 14)	*Proc. Natl. Acad. Sci. U.S.A.*	1795 (295 + 1500)	70 297
28 (37, 25)	*Izv. Vyssh. Uchebn. Zaved., Khim. Khim. Tekhnol.*	1660 (583 + 1077)	71 957
29 (53, 24)	*Biochem. Biophys. Res. Commun.*	1603 (504 + 1099)	73 560
30 (21, 45)	*Neftepererab. Neftekhim. (Moscow)*	1557 (762 + 795)	75 117
31 (15, 53)	*J. Electroanal. Chem. Interfacial Electrochem.*	1548 (852 + 696)	76 665
32 (38, 34)	*J. Chem. Soc., Faraday Trans. 1*	1494 (572 + 922)	78 159
33 (31, 43)	*Biochim. Biophys. Acta*	1476 (639 + 837)	79 635
34 (56, 30)	*J. Biol. Chem.*	1443 (471 + 972)	81 078
35 (28, 48)	*Zavod. Lab.*	1412 (670 + 742)	82 490
36 (51, 35)	*Chem. Phys.*	1404 (506 + 898)	83 894
37 (52, 38)	*Biochemistry*	1366 (505 + 861)	85 260

number of relevant papers and from the user's point of view are least essential.

The titles of journals at the top of the united list (joint results of volumes 86 and 89) are specified in Table II. Out of 9074 journals, the 37 top ones supplied 20% of the papers (compare Table I, column 4) selected for users from the *CACon* data base during the period of investigation.

The Findings As Compared with the Lists of Journals Most Frequently Cited in Chemical Abstracts. The lists of journals cited in SDI output (from volumes 86 and 89 of *CACon*) were compared with the lists of journals most frequently cited in *Chemical Abstracts*[26] in 1977 and 1978, and the following conclusion drawn.

(1) In the frequency list made up on the basis of references retrieved for SDI users from volume 86 of *CACon*, there were 986 common titles with the list of 1000 journals specified in the 1977 *CASSI* supplement, based on a coverage analysis of two volumes of *Chemical Abstracts*—86 and 87 (January–December 1977).

(2) In the list made up on the basis of references from volume 89 of *CACon*, there were 996 common titles as compared with the list of 1000 journals that was based on a coverage analysis of two volumes of *CA*—88 and 89 (January–December 1978).

(3) When a comparison between the titles of the 37 top journals (that delivered 20% of papers at the time of investigation) specified in Table II and the 37 top titles enumerated in the list of 1000 most frequently cited journals based on volumes 86 and 87 of CA and then the list based on volumes 88 and 89 of *CA* is drawn, it was found that the overlap of journal titles in this range was 21 in the first case and 20 in the second one.

(4) The ranks of biochemical journals are lower in the list compiled in the course of the present study (Table II, titles ranked 29th, 33rd, 34th, 37th) than in both the 1977 and 1978 lists of journals most frequently cited in *CA* (where they are ranked 12th and 10th, 1st and 1st, 4th and 5th, 17th and 23rd, respectively). This is probably an effect of concurrent use of MEDLARS and AGRIS, data bases also covering biochemistry.

As had been assumed, all primary journals (except a few examples of "nonchemical" titles and infrequently published periodicals, which were probably cited in only one of the two volumes published annually) specified in the *CASSI* lists of journals most frequently cited in *CA* were found in the references retrieved for users in this study. But the rank of titles, which in the *CASSI* lists is determined by the "productivity" of individual journals contributing to *Chemical Abstracts*, is modified in the frequency lists generated from this investigation by the users' subject area of interest, represented in profiles, which are included in the search procedure.

ACKNOWLEDGMENT

I thank Mr. Grzegorz Bogdan of the Chemical Information Center at Warsaw Technical University for his helpful cooperation, which made this study possible.

REFERENCES AND NOTES

(1) Williams, M. E. "The Impact of Machine-Readable Data Bases on Library and Information Services", *Inf. Process. Manage.* **1977**, *13*, 95–107.
(2) Schipma, P. B. "Generation and Uses of Machine-Readable Data Bases", *Annu. Rev. Inf. Sci. Technol.* **1975**, *10*, 237–271.
(3) Bourne, C. P. "Computer-Based Reference Services As an Alternative Means to Improve Resource Poor Local Libraries in Developing Countries", *Int. Libr. Rev.* **1977**, *9*, 43–50.
(4) Bourne, C. P. "Improvements in the Coupling of SDI System Output with Document Delivery Systems", *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 27–30.
(5) Hawkins, D. T. "Unconventional Uses of On-Line Information Retrieval Systems: On-Line Bibliometric Studies", *J. Am. Soc. Inf. Sci.* **1977**, *28*, 13–18.
(6) Hawkins, D. T. "Semiconductor Journals", *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 21–23.
(7) Kramárová, A. "Využívanie magnetickopáskovej služby INSPEC pri budovaní informačného fondu SLTK", *Technická Knihovna* **1977**, *21*, 329–332.
(8) Hoffmann, W. D.; Klawiter-Pommer, J. H. T. "Relevante Zeitschriften für die Veterinärmedizin—Versuch einer Analyse anhand von fünf Literatur-Datenbasen", *Nachr. Dok.* **1977**, *28*, 64–67.
(9) Singleton, A. "Journal Ranking and Selection: A Review in Physics", *J. Doc.* **1976**, *32*, 258–289.
(10) Subramanyam, K. "Criteria for Journal Selection", *Spec. Libr.* **1975**, *66*, 367–371.
(11) Scales, P. A. "Citation Analyses As Indicators of the Use of Serials: A Comparison of Ranked Title Lists Produced by Citation Counting and from Use Data", *J. Doc.* **1976**, *32*, 17–25.
(12) Rodgers, E. "Pruning periodical subscriptions at Glasgow University Library", *Aslib Proc.* **1978**, *30*, 145–153.
(13) Thrope, P. "Inter-Library Loans Analysis for Journals Acquisition", *Aslib Proc.* **1979**, *31*, 352–359.

(14) Wenger, C. B.; Childress, J. "Journal Evaluation in a Large Research Library", *J. Am. Soc. Inf. Sci.* **1977**, *28*, 293-299.
(15) Wender, R. W. "Counting Journal Title Usage in the Health Sciences", *Spec. Libr.* **1979**, *70*, 219-226.
(16) Garfield, E. "Citation Analysis As a Tool in Journal Evaluation", *Science (Washington, D.C.)* **1972**, *178*, 471-479.
(17) Cawkell, A. E. "Evaluating Scientific Journals with Journal Citation Reports—A Case Study in Acoustics", *J. Am. Soc. Inf. Sci.* **1978**, *29*, 41-46.
(18) Hirst, G. "Discipline Impact Factors: A Method for Determining Core Journal Lists", *J. Am. Soc. Inf. Sci.* **1978**, *29*, 171-172.
(19) Cezairliyan, A. O.; Lykoudis, P. S.; Touloukian, Y. S. "A New Method for the Search of Scientific Literature through Abstracting Journals", *J. Chem. Doc.* **1962**, *2*, 86-92.
(20) Windsor, D. A. "Rational Selection of Primary Journals for a Biomedical Research Library: The Use of Secondary Journal Citations", *Spec. Libr.* **1973**, *64*, 446-451.
(21) Pope, A. "Bradford's Law and the Periodical Literature of Information Science", *J. Am. Soc. Inf. Sci.* **1975**, *26*, 207-213.
(22) Danilowicz, C.; Szarski, H. "Metoda oceny wartości merytorycznej czasopism naukowych na podstawie danych systemu SDI", *Zagadnienia Nauk.* **1979**, *15*, 55-63.
(23) Stefaniak, B.; Kacprowicz, H.; Dabkowski, H.; Grzecznowska, A. "Selektywna Dystrybucja Informacji. Organizacja i Funkcjonowanie SDI w Polsce", *Prace IINTE* **1979**, *20*, 1-119.
(24) Stefaniak, B. "Investigation of Needs of Users of Scientific and Technical Information System in Chemistry for Information Sources Included in *CACon* Data Base", 1st International Seminar—Foreign Machine-Readable Data bases for Chemistry and Chemical Industry—Processing and Experience. Warsaw 1979, 65-77 (in Russian).
(25) Brookes, B. C. "Numerical Methods of Bibliographic Analysis", *Libr. Trends* **1973**, *22*, 18-43.
(26) The lists come from the 1977 and 1978 "Chemical Abstracts Service Source Index" (*CASSI*) cumulative supplements.

Registration–Identification of Crystalline Materials Based on Lattice and Empirical Formula

JOHN R. RODGERS†

University Chemical Laboratory, Cambridge CB2 1EW, England

ALAN D. MIGHELL*

Center for Materials Science, National Bureau of Standards, Washington, DC 20234

Received August 20, 1980

Data files containing information on solid-state materials are expanding rapidly. Each year several thousand new materials are characterized by X-ray diffraction techniques. Consequently, computer procedures have been developed to register materials entering large data files. We have found registration based on lattice parameters and empirical formula especially effective. In our present registration procedure, the lattice is uniquely represented by the reduced cell and the elements in the formula are uniquely specified by prime numbers. This method has been applied for several years to register new materials for the Cambridge Crystallographic Data File which contains data on more than 25 000 compounds containing organic carbon. The method is now being adapted to register materials for the NBS Crystal Data File. Our experience suggests the desirability of routinely characterizing organic materials by cell parameters in addition to the traditional chemical analysis. A solid-state registry number that would allow one to identify the same compound in different data bases could be derived from the lattice constants and the empirical formula. Such a number would make it possible to distinguish polymorphs and different phases of the same composition.

INTRODUCTION

The recent expansion in the number of crystal structures characterized by X-ray diffraction methods[1] has forced data-base builders to improve techniques for the registration of new materials. Registration is the process by which new entries are added (if acceptable) to a data base and includes the identification of existing entries in the master file that are the same as or related to a given new entry. During registration, the new entry is critically evaluated with respect to the existing entries, and it is assigned a registry number and a reference code. In some cases existing entries may be changed, updated, or deleted. The method of registration (lattice formula) described here has been used at the Cambridge Crystallographic Data Centre since 1977 and will soon be adapted as a general method for registration of new crystalline materials entering the NBS Crystal Data File.[2]

Registration based on lattice formula is an effective procedure to identify previous entries on the same material, isometrically and isostructurally related substances, and materials with related lattices (e.g., sub- and superlattices). For organic materials which crystallize mainly in the lower symmetry crystal systems,[3] our experience has shown that the lattice alone is highly characteristic for a given compound. In fact, it could be used as the basis of a solid-state registration number.

In lattice-formula registration discussed below, the lattice is specified by the reduced cell and the elements of the empirical formula are uniquely defined by prime numbers. The reduced cell is used because it can be uniquely defined[4-6] and convenient algorithms to calculate it[5,6] have been devised. For registration, this cell allows one to locate identical and metrically similar lattices independently of crystal system and lattice centering. Further reasons for using the reduced cell for identification of crystalline materials have been published.[7] Recently, more general matrix methods to establish interlattice relationships have been developed.[8] For registration enhancement and the critical evaluation of data, newer versions of the registration program will be expanded to include these matrix methods.

GENERAL METHODOLOGY OF LATTICE-FORMULA REGISTRATION

An overall view of the general techniques used in lattice-formula registration is given in Figure 1. The new data (EDIT FILE) are registered against the old data (MASTER FILE). This general procedure has been used by the Cambridge Crystallographic Data Centre to register a new batch of compounds before they are added to the MASTER FILE. A

†University of Adelaide, Adelaide, South Australia, 5001.