Cambridge Crystallographic Data Centre. I. Bibliographic File

OLGA KENNARD,* D. G. WATSON, and W. G. TOWN
Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England

Received November 29, 1971

The Cambridge Crystallographic Data Centre is concerned with the retrieval, evaluation, synthesis, and dissemination of structural data based on diffraction methods. This paper is Part I of a series describing the work of the Centre and deals with the organization of a computerized bibliographic file. Examples are given of the use of the file for bibliographic services, computer-typeset publications, and statistical analysis of trends in publications.

The Crystallographic Data Centre was established in 1965 with the financial support of the Office for Scientific and Technical Information (OSTI) as part of the British contribution to international collaboration on critical data compilations.[1,2] The Centre comes under the category of a "specialized data center" and as such it is concerned both with the organization of crystallographic information and basic crystallographic research. A system has been developed using an IBM 360/44, and the files contain both bibliographic information and numeric data on crystal and molecular structures determined by diffraction methods.

The Centre also provides input to the organic section of "Crystal Data."[3] This is a standard crystallographic reference publication which is being compiled at the National Bureau of Standards as part of the program of the National Standard Reference Data System.

SCOPE

The system is comprehensive for all carbon-containing compounds, excluding polymers and proteins, for which unit cell data have been determined by x-ray or neutron diffraction studies. Within this definition, two categories of crystallographic analysis are recognized:

(a) all organic and organometallic compounds whose structures have been fully determined—i.e., coordinates of each atom have been established
(b) all carbon compounds for which only unit cell data have been determined

Category (a) is the major concern of the Centre but both categories form the input to the "Crystal Data" project.

LITERATURE COVERAGE

The literature relating to organic and organometallic crystallography is not widely dispersed, being confined to less than 200 journals and a statistical analysis of recent accessions shows that seven journals (Table I) account for about 75% of the current publications.

The above journals are scanned in-house and Xerox copies of all relevant articles are made to maintain a complete reprint collection for the post-1959 literature. The reprint file is essential for the abstracting of data and, at a later stage, for referral purposes in the course of critical evaluation of the numeric data.

The remaining 25% of the literature is accessed by use of the abstracting journal, Bulletin Signalétique. The crystallography section (161) of this journal is ideally suited to our needs as special attention is given to the standardization of abstracts and to the controlled use of signals. Reprints are ordered from the Centre National de la Recherche Scientifique, Paris and standardized signals are important so that we can select appropriate articles on the basis of the abstracts.

The abstracts of proceedings of two major crystallographic conferences are regularly scanned as soon as they become available. These are the Conferences of the American Crystallographic Association and the Congresses of the International Union of Crystallography.

In the case of Russian publications, we receive periodic checklists from N. Smirnova of Moscow University. These lists are checked against the file and are useful for notification of articles published in minor journals. Wherever possible, we hold the English translations of the reprints, but in the bibliographic file, the page number of the original Russian publication is recorded.

BIBLIOGRAPHIC RECORD STRUCTURE

Bibliographic data abstracted from the original papers are recorded on file cards and also encoded for transfer to 80-column punched cards and finally to magnetic tape. The elements of the bibliographic record are:

Reference Code
Compound Name
Compound Synonym
Authors' Names
Journal Reference
Cross-Reference(s)
Molecular Formula
Chemical Classification

The file card carries, in addition to the above items, the chemical structural formula.

The coding of information for the computer file is based on a system of card types. Each card type is identified by the first six columns of the card—e.g., COMPND carries the compound name, FORMUL the molecular formula, etc. Details of the various bibliographic elements are given below.

* External Staff, Medical Research Council. To whom inquiries should be addressed.

Table I. Journals Publishing the Major Proportion of Crystallographic Papers

Acta Crystallographica
J. Chem. Soc.
J. Amer. Chem. Soc.
Inorg. Chem.
Acta Chem. Scand.
Zh. Strukt. Khim.
Tetrahedron Letters

Reference Code. Since nearly all crystallographic publications are concerned with reporting the structure of a single compound, the best way of "labeling" a paper is with respect to the compound name. Thus, the actual title of the paper can be ignored, and we can avoid the difficulties of retrieval based on titles.

We define an entry as a given publication on a given compound. Each entry is identified by its reference code: a six-letter plus two-digit code, the letters forming an acronym of the compound name. Thus a publication on the structure of benzene by Smith might have reference code BENZEN00 and the independent publication by Jones the reference code BENZEN01. In practice, the number of separate publications on the same compound is very small (less then four, except for some substances of great crystallographic interest where there might be a few more). If there is only one publication on the compound, we simply use the six-letter code—e.g. BENZEN. Further details of the reference code system will be discussed in the section dealing with registration.

Compound Name and Synonym. If the compound name used by the author is correct, then this is recorded and, where appropriate, a synonym might be introduced. In the case of faulty or ambiguous names, a correct systematic name is assigned.

For identification purposes, or for defining some aspect of the paper, we allow the use of qualifying phrases following the compound name. If no such phrase is present then the implication is that the compound was studied at room temperature by x-ray diffraction, and the author has not reported the existence of polymorphism. Examples of qualifying phrases are: at −150°C, neutron study, and tetragonal form.

When the file was first established, consideration was given to the problem of computer typesetting. For our purposes, the typesetting of compound names involves handling the following: upper- and lower-case alphabetics, superior and inferior alphabetics and numerals, Greek characters, and miscellaneous symbols—e.g., ±, °, etc.

The IBM 026 card punch was available to us, and a system was devised based on the BCD character set—viz., upper-case alphabetics, numerals, and 11 miscellaneous symbols. Details of the scheme which has been successfully implemented will be given in a later paper in this series.

Authors' Names. The full list of authors' names is recorded, and a simple signalling system is used to assist in the typesetting of multiword surnames and other names which cannot be easily handled by syntax analysis.

Journal Reference. The journal reference takes the usual form: name or abbreviation, volume, page, year. However, a coden is also assigned to each journal or part-journal, and this enables us to typeset the journal name by a simple table look-up.

Cross References. For literature prior to 1960, cross-references are provided to standard reference works on crystal and molecular structures. These are:

(i) "Structure Reports"[4] a continuation of "Strukturbericht," since 1940, published by the International Union of Crystallography (IUCr)

(ii) "Tables of Interatomic Distances and Configuration of Molecules and Ions," Chemical Society Special Publications, No. 11 (1958) and No. 18 (1965).

For (i) we record the volume, page, year and for (ii) the page of the main or supplement volumes.

Molecular Formula. The bibliographic file is linked, through the reference code, to the structural data file which contains—among other numeric data—the atomic coordinates which, for any given structure, define the bonded network of atoms. For this reason, we have chosen to express the molecular formula in terms of residues, each residue being an independent set of bonded atoms. Two examples will serve to illustrate the scheme:

(i) sodium acetate monohydrate consists of three residues (R):

R1 is $C_2H_3O_2^-$
R2 is Na^+
R3 is H_2O

(ii) the molecular complex between azulene and s-trinitrobenzene has two residues:

R1 is $C_{10}H_8$
R2 is $C_6H_3N_3O_6$

The molecular formula is coded on FORMUL cards, one residue per card, using the element sequence: C, H, other elements in alphabetic order. Provision is made for charges and pre- and post-multipliers. Typesetting of molecular formulas is achieved by syntax analysis.

Chemical Classification. At the outset of the project it was recognized that some type of classification scheme would be needed to assist the searching of the file. Compound name indexes involve problems of alternative nomenclature, and molecular formula indexes involve the isomer problem, even in a relatively small file of less than 10,000 compounds. Furthermore, in structural chemistry, the research worker is usually interested not just in locating the reference to a single compound but also in retrieving information on similar chemical structures. For these reasons a classification scheme based on chemical structural type was devised. The total list of classes is given in Table II.

It should be stated that certain classes have been recognized because of intensive crystallographic work in these areas—thus barbiturates are separated from other pyrimidines and naphthoquinones from other naphthalenes, etc. Also the important decision was made to classify coordination complexes and π-complexes in terms of the ligand rather than the transition metal. Class 50 (antibiotics) is the only class which is functional rather than structural. The classes of terpenes, alkaloids, etc., are, of course, not strictly structural but, nevertheless, well suited for retrieval purposes.

The classification scheme is based on the concept of residues described above. Only organic residues are classified, and each compound is assigned to a basic class with provision for up to eight cross-references to other classes. A few examples will illustrate the procedure:

1. Deoxyadenosine monohydrate
   Two residues but only R1 is organic.
   Basic class: 47 (nucleosides and nucleotides)
   Cross references: 44 (pyrimidines and purines)
   45 (carbohydrates)
2. π-Allyl-π-cyclopentadienyl palladium
   Basic class: 72 (metal π-complexes, open chain)
   Cross references: 73 (metal π-complexes, cyclopentadiene)
   In this case, 73 could have been chosen as the basic class
3. Molecular complex of azulene and s-trinitrobenzene
   Two residues: R1 is $C_{10}H_8$ and R2 is $C_6H_3N_3O_6$
   Basic class: R1 in 60 (molecular complexes)

Table II. List of Chemical Classes

Class No.		Class No.		Class No.	
1	Aliphatic Carboxylic Acid Derivatives	29	Hydrocarbons (4 Fused Rings)	60	Molecular Complexes
2	Aliphatic Carboxylic Acid Salts (Ammonium, IA, IIA Metals)	30	Hydrocarbons (5 or More Fused Rings)	61	Clathrates
3	Aliphatic Amines	31	Bridged Ring Hydrocarbons	62	Boron Compounds
4	Aliphatic (N and S) Compounds	32	Hetero-Nitrogen (3,4,5-Membered Monocyclic)	63	Silicon Compounds
5	Aliphatic Miscellaneous			64	Phosphorus Compounds
6	Enolates (Aliphatic and Aromatic)	33	Hetero-Nitrogen (6-Membered Monocyclic)	65	Arsenic Compounds
7	Nitriles (Aliphatic and Aromatic)	34	Hetero-Nitrogen (7- and Higher-Membered Monocyclic)	66	Antimony and Bismuth Compounds
8	Urea Compounds (Aliphatic and Aromatic)	35	Hetero-Nitrogen (2 Fused Rings)	67	Groups IA and IIA Compounds
9	Nitrogen-Nitrogen Compounds (Aliphatic and Aromatic)	36	Hetero-Nitrogen (More than 2 Fused Rings)	68	Group III Compounds
10	Nitrogen-Oxygen Compounds (Aliphatic and Aromatic)	37	Hetero-Nitrogen (Bridged Ring Systems)	69	Germanium, Tin, Lead Compounds
11	Sulphur and Selenium Compounds	38	Hetero-Oxygen	70	Tellurium Compounds
12	Carbonium Ions, Carbanions, Radicals	39	Hetero-Sulphur and Hetero-Selenium	71	Transition Metal-C Compounds
13	Benzoic Acid Derivatives	40	Hetero- (Nitrogen and Oxygen)	72	Metal π-Complexes (Open-Chain)
14	Benzoic Acid Salts (Ammonium IA, IIA Metals)	41	Hetero- (Nitrogen and Sulphur)	73	Metal π-Complexes (Cyclopentadiene)
15	Benzene Nitro Compounds	42	Hetero-Mixed Miscellaneous	74	Metal π-Complexes (Arene)
16	Anilines	43	Barbiturates	75	Metal π-Complexes (Miscellaneous Ring Systems
17	Phenols and Ethers	44	Pyrimidines and Purines	76	Metal Complexes (Ethylenediamine)
18	Benzoquinones	45	Carbohydrates	77	Metal Complexes (Acetylacetone)
19	Benzene Miscellaneous	46	Phosphates	78	Metal Complexes (Salicylic Derivatives)
20	Monocyclic Hydrocarbons (3,4,5-Membered Rings)	47	Nucleosides and Nucleotides	79	Metal Complexes (Thiourea)
21	Monocyclic Hydrocarbons (6-Membered Rings)	48	Amino-Acids and Peptides	80	Metal Complexes (Thiocarbamate or Xanthate)
22	Monocyclic Hydrocarbons (7,8-Membered Rings)	49	Porphyrins and Corrins	81	Metal Complexes (Carboxylic Acid)
23	Monocyclic Hydrocarbons (9- and Higher-Membered Rings)	50	Antibiotics	82	Metal Complexes (Amino-Acid)
24	Naphthalene Compounds	51	Steroids	83	Metal Complexes (Nitrogen Ligand)
25	Napthoquinones	52	Monoterpenes	84	Metal Complexes (Oxygen Ligand)
26	Anthracene Compounds	53	Sesquiterpenes	85	Metal Complexes (Sulphur or Selenium Ligand)
27	Hydrocarbons (2 Fused Rings)	54	Diterpenes	86	Metal Complexes (Phosphine or Arsine Ligand)
28	Hydrocarbons (3 Fused Rings)	55	Sesterterpenes		
		56	Triterpenes		
		57	Tetraterpenes		
		58	Alkaloids		
		59	Miscellaneous Natural Products		

Cross references: R1 in 27 (hydrocarbon, 2 fused rings)
R2 in 60 (molecular complexes)
R2 in 15 (benzene nitro compounds)

Entries within a given class are ordered by increasing molecular formulas. Thus, for Example 3, the main entry would occur in class 60 at the $C_{10}H_8$ position, and there would be three cross-reference entries as follows:

in class 27 at $C_{10}H_8$
in class 60 at $C_6H_3N_3O_6$
in class 15 at $C_6H_3N_3O_6$

In this way, someone interested in all molecular complexes of azulene need only look at the $C_{10}H_8$ section of class 60. Alternatively, a research worker who is interested in azulene compounds would presumably search first of all in class 27 and might not think of molecular complexes. However, because of the cross-referencing system, he will find references to these in class 27, and thus the recall power of the search is enhanced.

REGISTRATION PROCESS

At present a new batch of about 180 entries is processed every 6 weeks. The registration process is carried out manually and is shown schematically in Figure 1.

The molecular formula of the new entry is compared against the molecular formula index for the total file resulting in no-hit or hit.

No-hit: The entry corresponds to a new compound, and a suitable reference code ABCDEF is assigned. This code must be checked against the alphabetic reference code index for the total file to avoid duplication.

Hit: In this case, the new entry belongs to one of three categories

(i) entry for a new compound
(ii) entry which supersedes an existing one—e.g., full publication following a preliminary communication by the same author
(iii) entry which supplements an existing one—e.g., publication by different authors

Category (i) is easily recognized by comparing the names and/or the structural formulas of the two isomers and a six-letter code assigned as for the "no-hit" case.

Examination of the reprints will differentiate categories (ii) and (iii).

(ii) If the existing entry has reference code ABCDEF, then the new entry will have reference code ABCDEF10. Likewise, ABCDEF11 supersedes ABCDEF01, etc.

(iii) If the existing entry has reference code ABCDEF then the new entry will have reference code ABCDEF01. If the existing entries are ABCDEF and ABCDEF01, then the new entry will have reference code ABCDEF02, etc.

When the registration process is complete the bibliographic record is coded and the cards punched, all cards for a given entry carrying the reference code in columns 73–80.

FILE UPDATING

The stages involved in the course of file updating for a new batch are:

(a) Entry Composition
(b) Checking Process
(c) Merging
(d) Index Production
(e) Statistics Compilation

Entry Composition. Cards are punched by a bureau operation, and the most efficient service has been obtained by punching all COMPND cards, followed by all AUTHOR cards, etc., rather than punching all information for a given entry before proceeding to the next. Accordingly, program BATSORT assembles the cards into entries on the basis of reference codes and sorts the entries alphabetically on these codes.

Checking Process. For routine purposes, three programs are used to check the information:

(i) GRAPHCHK examines all punched characters and determines whether each character corresponds to a printable graphic symbol.

(ii) BIBCHECK is the principal check program. It deals with a multitude of problems including syntax checking, field validity, format errors, missing cards, etc. Further details will be given in the paper on typesetting.

(iii) BIBLST2 is used to check the typesetting signals for compound names, synonyms, and authors' names. It lists each entry using a 120-character line-printer chain, and the output is scanned for errors.

Merging. When all the errors have been corrected, the batch of new entries is merged with the master bibliographic file, each entry being slotted in its basic class at the correct molecular formula position. This program BIBMERGE also takes account of new entries which supersede existing entries. Thus, if a new entry has reference code ABCDEF10 then the entry for ABCDEF on the master file will be deleted automatically and this action logged. The new master file is listed after each update.

Index Production. After merging a new batch various indexes are produced:

(i) molecular formula
(ii) reference code
(iii) journal references by coden and chronologically for each coden

Statistics Compilation. Program BIBSTAT is run against each new master file, and miscellaneous statistics are compiled. Some of these, for the current master tape, are presented in the statistics section.

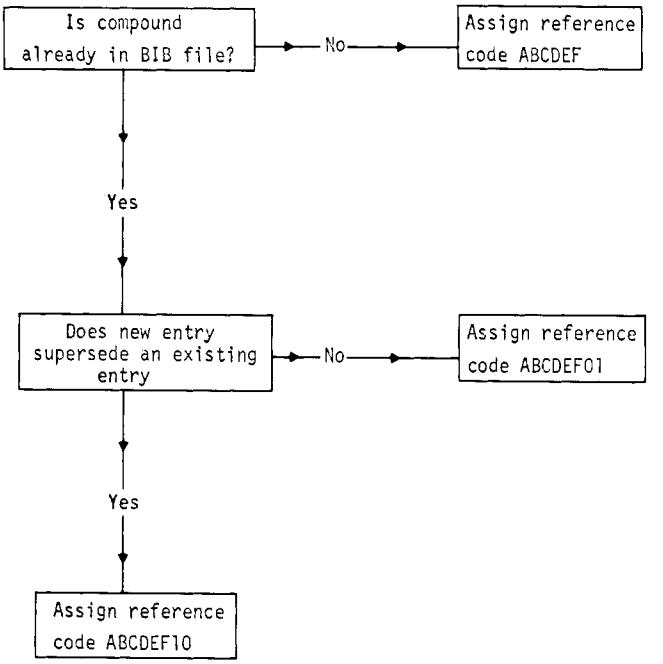

Figure 1

SERVICES

The principal bibliographic services provided by the Centre fall into three categories:

(a) "Molecular Structures and Dimensions"[5]
(b) "Structure Reports"
(c) Class Lists

"Molecular Structures and Dimensions." This is a new series of standard reference volumes published by the Centre in conjunction with the IUCr. The series will include both bibliographic issues and numeric data compilations. The first two volumes were published in 1970 and constituted bibliographies for the period 1935–1969.

Volume 1 covers general organic crystal structures, and volume 2 covers complexes and organometallic structures. The entries are classified as discussed earlier, and the two volumes contain information on some 4000 structures. Molecular formula, author, and transition metal indexes are provided. The volumes are typeset by Unwin Brothers Limited using the Photon 713 filmsetter and reproduced by photo-offset. The input magnetic tapes for the Photon were generated from our file using programs developed by the Centre and by INSPEC (Information Service in Physics, Electrotechnology and Computers and Control). Specimen entries are shown in Figure 2.

The first supplement to the bibliography, covering 1970/71 literature, will be published early in 1972 and will contain cumulative indexes. It is planned that after three or four supplements, a new edition will be produced, and at that time, we can reconsider classification problems resulting from user feedback.

"Structure Reports." This series forms a continuation of "Strukturbericht" which covered the literature to 1939 and is now published by the International Union of Crystallography. Each volume provides reports on all crystal structures for a given year. To each organic editor, we supply a bibliographic list for the specific year, and we undertake to notify him of any later publications which supplement or supersede entries for that year. The Commission on Structure Reports of the IUCr has formerly adopted our classification scheme and will incorporate it from the 1963 volume onwards.

15.20 **o - Nitrobenzaldehyde (stable form)**
$C_7H_5NO_3$
P.Coppens, G.M.J.Schmidt *Acta Cryst.*, **17**, 222, 1964

15.21 **o - Nitrobenzaldehyde (neutron diffraction)**
$C_7H_5NO_3$
P.Coppens *Acta Cryst.*, **17**, 573, 1964

41.C **Gliotoxin (at −150 ° C, absolute configuration)**
$C_{13}H_{14}N_2O_4S_2$
For complete entry see 50.7

51.12 **5α - Bromo - 6β,19 - oxido - pregnan - 3β - ol - 20 - one**
$C_{21}H_{31}BrO_3$
E.M.Gopalakrishna, A.Cooper, D.A.Norton
Acta Cryst. (B), **25**, 2473, 1969

60.118 **Tricarbonyl chromium anisole - 1,3,5 - trinitrobenzene complex**
$C_{10}H_8CrO_4$, $C_6H_3N_3O_6$
O.L.Carter, A.T.McPhail, G.A.Sim *J. Chem. Soc. (A)*, 822, 1966
Residue 1 also classified in 74; residue 2 classified in 60, 15

73.17 **Dicyclopentadiene platinum chloride**
$C_{10}H_{10}Cl_2Pt$
N.C.Baenziger, J.R.Doyle, G.F.Richards, C.L.Carpenter
Adv. Chem. Coord. Compounds,
Proc. 6th Internat. Conf., Detroit, 131, 1961

Figure 2

Table III. Growth of Structure Analysis by Diffraction Methods

(Number of entries per year)

Year	Entries	Year	Entries
Pre-1960	532	1965	407
1960	119	1966	542
1961	140	1967	674
1962	156	1968	781
1963	249	1969	1091
1964	253	1970	1185
		1971	427

Table IV. Distribution by Year among Major Class Groupings

(Values are percentages of total file)

Year	I[a]	II[b]	III[c]	IV[d]
Pre-1960	60	18	7	15
1960	48	12	13	27
1961	58	11	13	18
1962	40	24	15	21
1963	43	21	14	22
1964	41	20	14	25
1965	43	18	12	27
1966	38	16	14	32
1967	39	14	14	33
1968	32	19	13	36
1969	32	19	12	37
1970*	31	19	12	38
1971*	34	19	10	37

[a]"Simple" organic (1-42).
[b]Natural products (43-59).
[c]Mol. complexes, organometals, organometalloids (60-70).
[d]π-complexes and coordination complexes (71-86).

Class Lists. The majority of queries addressed to the system have been capable of solution with respect to the chemical classification scheme. Typically, a research worker wants all information relating to compounds of a certain structural type, and by specifying the appropriate class numbers, the relevant subset of the file can be listed. This type of service has proved of immense value not only to bench scientists but also to the authors of review articles.

Those queries which could not be answered by this technique have involved the search for small functional groups. The only satisfactory method of handling such problems would be in terms of fragment codes, linear notations, or connection tables.

STATISTICS

In discussing statistics relating to the bibliographic file, values for the years 1970/71 are always asterisked. This is to remind readers that data are still incomplete for 1970/71. Also it should be borne in mind that, as described earlier, entries can be superseded at a later date, and no account is taken of superseded entries in these statistics.

The total number of entries is 6556 and the total number of compounds 6012. The difference, 544, represents multiple studies of the same compounds. The growth of the subject is shown in Table III.

Table IV gives an analysis of the distribution of entries by year among major class groupings.

Whereas the study of simple organic structures has diminished, that of π-complexes and coordination complexes has greatly increased. This is presumably a result of the growing industrial importance of complexes and it also may reflect the investment of considerable governmental funds in this area of chemistry. The pronounced upsurge of activity in natural products about 1962 is possibly due to the fact that at that time computers were becoming more generally available in universities, thus allowing the solution of these fairly large structures.

As indicated earlier, seven journals currently account for about 75% of the literature in this field, and Table V shows the contribution of these journals.

There appears to be a relative decrease in the use of *Acta Cryst.* with a corresponding increase in the role of *J. Chem. Soc.* and *J. Amer. Chem. Soc.* During the past few years, *Inorg. Chem.* has made a marked impact and has established a considerable reputation in the field of metal com-

Table V. Distribution by Year among Major Journals

(Values are percentages of total file)

Year	I[a]	II[b]	III[c]	IV[d]	V[e]	VI[f]	VII[g]
Pre-1960	50	10	6	—	4	—	0
1960	47	7	1	—	5	0	1
1961	48	13	1	—	4	3	0
1962	46	15	3	0	8	2	0
1963	43	12	4	4	3	2	1
1964	37	10	8	8	5	3	0
1965	40	12	6	5	5	2	1
1966	38	16	6	7	5	3	1
1967	23	22	7	7	5	3	1
1968	21	23	9	8	5	4	3
1969	31	21	8	7	4	3	3
1970*	17	31	10	6	4	2	3
1971*	28	45	2	8	1	0	1

[a]*Acta Crystallographica.* [b]*J. Chem. Soc.* [c]*J. Amer. Chem. Soc.* [d]*Inorg. Chem.* [e]*Acta Chem. Scand.* [f]*Zh. Strukt. Khim.* [g]*Tetrahedron Letters*—indicates nonexistence of journal for stated year.

plexes. The relative swing from *Acta Cryst.* to chemical journals is presumably influenced by a recent change in the nature of organic crystallography. Nowadays, fast computers and powerful structure-solving techniques have created a situation where many crystal structure analyses are only part of a more comprehensive study. Thus, the emphasis is shifting from the crystallographic to the chemical nature of the problem.

FUTURE DEVELOPMENTS

The development projects in which the Centre is actively engaged come under three categories:

(a) Input
(b) Substructure Search
(c) Data Base Extension

Input. From economic and other considerations, it would be advantageous if optical character recognition (OCR) could be utilized for input to the files. Trials have been conducted using Scan Data hardware to read bibliographic records, and these tests have been highly satisfactory. Over the next few months, it is intended to input material both by OCR and punched cards and switch entirely to the former as soon as the system is fully tested.

Substructure Search. As indicated earlier, certain search queries are not easily answered in terms of our classification scheme, and a deeper search capability is obviously needed. To this end, we are at present examining the possibility of encoding compounds in the Wiswesser Line Notation (WLN) and in the form of connection tables. Because of the wide range of compounds in our file, the WLN approach presents many problems in areas where rules have not yet been formalized.

Data Base Extension. The data base is being extended to cover molecular structures determined in the gas phase. A file is currently being set up to include structures studied by microwave and gas-phase electron diffraction studies. The first output from this file will be a bibliography, and it is planned to publish this as part of the series "Molecular Structures and Dimensions."

The data base is also being extended to protein structures, and the establishment of a Protein Data Bank has just been announced.[6]

ACKNOWLEDGMENT

We thank the Office for Scientific and Technical Information for their support of the activities of the Crystallographic Data Centre and individual staff members for their advice and helpful discussion. Space for the Centre was provided by Cambridge University and computing facilities by the Institute of Theoretical Astronomy. We are grateful to INSPEC and especially to P. Simmons for the use of their computer typesetting program. We thank R. S. Cahn and the Nomenclature Division of the Chemical Abstracts Service for advice on nomenclature, also P. Baker and his colleagues at ICI Pharmaceuticals Division for their assistance with the WLN system.

LITERATURE CITED

(1) Hall, R. M. S., *J. Chem. Doc.*, **1**, 18 (1967).
(2) "CODATA International Compendium of Numerical Data Projects," Springer-Verlag, Berlin, 1969.
(3) Donnay, J. D. H., Donnay, G., Cox, E. G., Kennard, O., and King, M. V., "Crystal Data, Determinative Tables, Part II," 2nd. ed. Monograph 5 of the American Crystallographic Association, Williams & Heintz, Washington, D. C., 1963.
(4) "Structure Reports," Vols. 8–24, published for the International Union of Crystallography by N. V. A. Oosthoek's Uitgevers Mij, Utrecht.
(5) Kennard, O., and Watson, D. G., "Molecular Structures and Dimensions," Vols. 1 and 2, published for the Crystallographic Data Centre, Cambridge, and the International Union of Crystallography by N. V. A. Oosthoek's Uitgevers Mij, Utrecht, 1970.
(6) *Nature New Biology* **233,** 223 (1971).

SSIE—An Information Center Which Stores Foresight*

FRANK J. KREYSA
Smithsonian Science Information Exchange, Inc.,
1730 M Street, N.W., Washington, D.C., 20036

Received November 11, 1971

In June 1971, the Smithsonian Science Information Exchange (SSIE) adopted the fourth official name it has borne in its 22-year history. These changes in name have been made to reflect expansions in the Exchange's coverage of subject matter or changes in administrative or fiscal structure. During this same period, the Exchange's methods of operation have undergone improvement and modernization, and it now has a much wider spectrum of services to offer the scientific and engineering community than ever before. The new, nonprofit organizational structure now in effect is shown in Figure 1.

*Presented at the American Chemical Society Northeast Regional Meeting, Buffalo, N. Y., October 12, 1971.

Since its founding in 1949, the SSIE and its predecessor organizations have been working in a unique field of information dissemination. The primary mission of the Exchange is to facilitate more effective planning, management, and coordination of the scientific research and development sponsored or supported by U.S. Government agencies. This is done by:

> Developing and maintaining an up-to-date, comprehensive inventory of all types of ongoing, unclassified research projects.
>
> Providing means and services to assist research and development administrators and other research and development personnel in obtaining and using the information stored.