19.X-05 THE NBS CRYSTAL DATA FILE. By Alan D. Mighell and Judith K. Stalick, Center for Materials Science, National Bureau of Standards, Washington, DC 20234.

The NBS Crystal Data File is designed for the characterization of solid-state materials. As the lattice parameters are highly diagnostic for a given compound, the file in conjunction with modern diffractometry offers an excellent way to identify unknown substances.

The file contains chemical and crystallographic data on approximately 60,000 substances of all types. It is expanding at the rate of about 5,000 compounds per year. For each compound, the following data items are included: the chemical name and formula, density, reduced cell, space group, space group number, and the literature reference. The data are obtained from the original literature by the NBS Crystal Data Center, by independent abstractors, and by the following cooperating data centers: Cambridge Crystallographic Data Centre, Inorganic Structural Data Center, Metals Data Center, and the JCPDS--International Centre for Diffraction Data.

To systematically transform and evaluate the entries for the file, a program called NBS*AIDS80 (available for distribution) has been developed. Program functions include determination of standard cells and space groups, assignment of space group numbers, determination of metric symmetry and checking of the data for consistency. In applying the program, it has been noted that full structure determinations are sometimes inadvertently carried out more than once, that symmetry (especially rhombohedral) is often missed, and that cells are sometimes not transformed correctly. Routine use of the program to analyze the data prior to publication would help to reduce such errors.

The data in the file are made available to the scientific community in several ways. The principal publication is Crystal Data Determinative Tables, in which substances are ordered by crystal system and unit cell ratios. This arrangement groups isostructural materials together. A new publication, Crystallochemical Identification Tables, is being prepared in which the data are sorted on the reduced cell parameters and like lattices are grouped together. To identify an unknown, one only needs to determine a primitive cell of the lattice, reduce this cell, and check the tables for a match. In addition to printed publications, a computer file on magnetic tape may be purchased from the NBS Office of Standard Reference Data; the file may also be searched on line through the Chemical Information System.

Projected future applications of the file include: 1) extensive use of the file in automated identification systems in which both cell determination and the file search are under computer control; 2) wider use of the file by the chemical community for the routine characterization of materials; 3) use of the file in conjunction with the powder diffraction file; 4) research projects carried out on the data within the file. Many lattice related research projects can be undertaken. One can systematically analyze such structural and lattice-related phenomena as: symmetry, coincidence-site lattices, derivative lattices, space group frequencies, and twinning. Two research projects that demonstrate the basic importance of the metric lattice in crystallography will be discussed.

19.X-06 CAMBRIDGE CRYSTALLOGRAPHIC DATABASE. Olga Kennard, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.

The Cambridge Crystallographic Database is well known to structural crystallographers and is available through Affiliated Data Centres in some 21 countries. It can be searched by a variety of program systems; a recent advance is a mini-computer version of the Cambridge system developed by R. E. Rosenfield for a PDP11/45. This opens up the possibility of direct use of the database on some of the larger diffractometer systems, which will be of importance in utilizing methods of structure solution requiring accurate molecular geometry of substructural fragments.

The past few years have seen an increasing use of the database, primarily for reviews and research projects. The database itself has been upgraded and the most recent release (January 1981) contains only 3% error-sets. There are, however, a substantial number of entries without numeric data. The question of even closer interaction with the scientific community and Journal Editors, with a view to maintaining and improving database content, will be discussed.

The major technical innovation during the past year has been the introduction of digitized chemical structural diagrams, with implications for new types of printed publications. Future plans include the creation of a new 'search' file and the views of the audience will be sought on the proposed contents of such a file, and on any new innovations which might improve database utility.

19.X-07 THE METAL DATA FILE. L.D. Calvert. Chemistry Division, National Research Council of Canada, Ottawa, Canada, K1A 0R9.

The Metal Data File (MDF) is a computer-based file in the Chemistry Division of the National Research Council of Canada. It contains the relevant data for the ~4800 intermetallic compounds with structures which have been refined in the period 1913-1979; 1980 data are being compiled and processed. In addition assigned structures amounting to a total of about 2900 have been recorded since 1975. The file is resident on a PDP-8e and is operated by programs written in Fortran. The file structure is similar to that of the Cambridge Crystallographic Database. Strong emphasis has been placed on data integrity and evaluation. Input is kept simple by using the general space group interpreting routine of A.C. Larson (Acta Cryst. (1969), A25, S1) which operates directly on the Hermann-Mauguin symbol and by using free format, numerical data input which is later formatted under program control. Data are verified by diagnostic programs which check for transcription errors and for data inconsistencies. The formula is computed from the structural parameters and compared with that given by the authors; the cell volume is compared with that expected for metallic cubic close packing; the density and volume per atom are also calculated. The major check is by interatomic distances which are selected by the criterion of Frank and Kasper (Acta Cryst. (1958) 11, 184), with appropriate allowance for atomic radii and cell volumes, and then calculated and compared with the sum of atomic radii as well as the distances given by the authors; unusual values are flagged. It has been found that up to about 25% of the original papers contain numerical errors, most of which are quite minor, but a few percent are significant. Most of these can be corrected either by an analysis of the data given or by

correspondence with the authors. In a few cases the problems are not resolved and such entries are either flagged or omitted.

The MDF is used to check the data appearing in the Metals section of Structure Reports and also to calculate peak intensity (i.e. full profile) powder patterns for the Powder Data File (PDF) published by the Joint Committee on Powder Diffraction Standards. In addition the relevant current data on metals are sent to Crystal Data. It is planned that patterns now missing from the PDF or those requiring checking will be calculated from the structural data when they are available.

In addition to the major routines which calculate interatomic distances and powder patterns there are also routines which can display the Frank-Kasper domains, the coordination polyhedra for individual atoms or the bonds between an atom and its neighbours or combinations of these. The displays can be produced in stereo, and rotated and copied onto the line-printer. There are also search routines to find specific compounds or compounds containing various elements or combinations of elements. Indices sorted by elemental type or by structure code can be prepared and are routinely used for searching the file.

At present the file is available for private use only but a joint program is now underway to make the file publicly available through the facilities of the Canada Institute for Scientific and Technical Information (CISTI). The file in its present form will be demonstrated at the Congress.

19.X-08 THE PROTEIN DATA BANK*. By Thomas F. Koetzle, Enrique E. Abola and Frances C. Bernstein, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973, USA

The Protein Data Bank (F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, J. Mol. Biol. (1977) 112, 535) is a computer-based archival file for structural data on biological macromolecules. The Bank presently contains atomic coordinates for more than 100 distinct molecules; RNA, DNA and polysaccharide structures are represented as well as proteins. Structure factors and phase information have been deposited for some of these studies. Bibliographic citations are included for structures when coordinates are not yet available. Data are distributed upon request from Brookhaven and centers in Cambridge, England; Osaka, Japan; and Clayton, Victoria Australia. A number of portable computer programs have been developed to enable users to extract useful information from the file.

*Work supported by the U. S. National Science Foundation under grant PCM 80 13028 and carried out under the auspices of the U. S. Department of Energy, Office of Basic Energy Sciences.

19.X-09 JCPDS--INTERNATIONAL CENTRE FOR DIFFRACTION DATA. By C. R. Hubbard, National Bureau of Standards, Washington, DC, USA and G. J. McCarthy, North Dakota State University, Fargo, ND, USA.

The JCPDS--International Centre for Diffraction (ICDD) is the publisher of the Powder Diffraction File (PDF). The PDF is used for identification and characterization of crystalline phases by reference to the characteristic x-ray powder diffraction pattern. The JCPDS--ICDD consists of a permanent staff based at Swarthmore, PA, and at the National Bureau of Standards, plus a voluntary organization of internationally known scientists who act as advisors. The PDF is published in a wide variety of forms and in several subsets. The primary product, a compendium of powder diffraction patterns tabulated in numerical form, is published yearly on 3" x 5" index cards in sets of approximately 2000 patterns. Today nearly 35000 cards comprise the entire PDF. The card images are also published on microfiche and in books. These card images contain (1) compound name and formula, (2) numerical pattern data, (3) a quality mark, (4) the PDF code number, (5) reference, (6) cell parameters, space group and density, (7) method of data acquisition, (8) other information characterizing the compounds, and (9) editorial comments. Along with the data cards three search manuals are published which enable a user to locate a pattern. They are the Alphabetical Index, the Hanawalt Method Search Manual and the Fink Method Search Manual.

Because of the size of the PDF, the data base has been divided into Inorganic and Organic & Organometallic subsets to assist in identification procedures and to serve the needs of specialized groups. In addition, subfiles of Minerals, Metals and Alloys, Common Phases, and NBS Patterns exist. These subfiles, except for the Common Phases, are published as card images in bound books along with appropriate search manuals. The Common Phases subfile is published only as a search manual supplement to the PDF itself.

An increasingly important method to assist in identification is computer search/match procedures. The JCPDS leases several computer readable forms of the file containing data items (1) - (4), for use with a variety of maxi and mini computer systems. The PDF can also be searched on two time sharing networks: Diffraction Data Tele Search of University Computing Co. and the Powder Diffraction Search Match (PDSM) component of the Chemical Information System (CIS). CIS provides cross linking between a large number of other data bases covering such areas as toxicology, mass spectra, Crystal Data, and Molecular Structure and Dimensions. The introduction of automated powder diffractometer systems incorporating the full file on magnetic disks is a major development in methods to assist users of the file.

Computer aided editorial evaluation is now being used to upgrade the reliability and quality of data entering the file. The principal program for this evaluation and data base building is NBS*AIDS80. The JCPDS--ICDD is extending their use of this program to evaluate all the data entering the file. Methods to reevaluate and, if necessary, correct previously published data are being explored. The evaluation of an entry entering the file includes those tests performed for a Crystal Data entry plus tests unique to indexed powder patterns. Included are checks on systematic absences, agreement between observed and calculated two-theta, calculation of M(20) and F(N) figures-of-merit, (de Wolff, J. Appl. Cryst., 1, 108, 1968; Smith and Snyder, J. Appl. Cryst., 12, 60, 1979) and a test for the presence of systematic errors as a function of two-theta (Young, Ann. Math. Stat., Vol. XII, 293, 1941).