Acta Crystallographica Section B
**Structural Science**
ISSN 0108-7681

# The Powder Diffraction File: present and future

**John Faber* and Tim Fawcett**

International Centre for Diffraction Data, Newtown Square, PA 19073, USA

Correspondence e-mail: faber@icdd.com

Received 1 February 2002
Accepted 18 February 2002

The International Centre for Diffraction Data (ICDD) produces the Powder Diffraction File (PDF). This paper discusses some of the seminal events in the history of producing this primary reference for powder diffraction. Recent key events that center on collaborative initiatives have led to an enormous jump in entry population for the PDF. Collective efforts to editorialize the PDF are ongoing and provide enormous added value to the file. Recently, the ICDD has created a new series of the PDF, designated PDF-4. These relational database structures are being used to house the PDF of the future. The design and benefits of the PDF-4 are described.

## 1. Introduction

The PDF has been the primary reference for powder diffraction data for over 50 years. Although the first publication appeared in 1938, the importance of the diffraction information was central to the formation of the Joint Committee for Chemical Analysis by Powder Diffraction Methods, co-sponsored by ASTM Committee E-4, the Crystallographic Society of America and the British Institute of Physics. The first reprint appeared in 1941 as Set 1 of the PDF. Since this date, a total of 51 Sets containing experimental data and 19 Sets of calculated patterns have been published. Combined, there have been over 130000 published patterns. A detailed historical account has been published by Smith & Jenkins (1996).

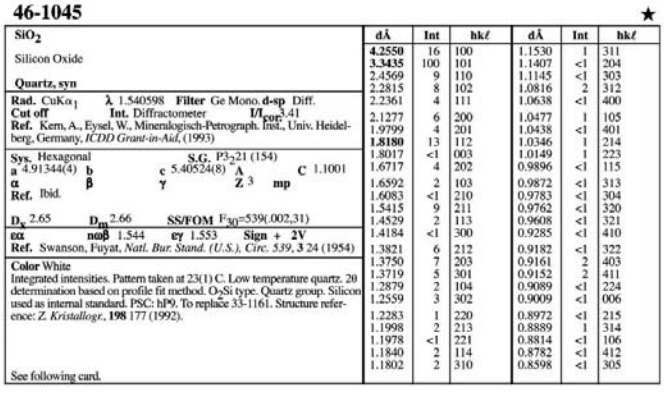
The primary information in the PDF is the collection of $d-I$ data pairs, where the $d$-spacing ($d$) is determined from the angle of diffraction, and the peak intensity ($I$) is obtained experimentally under the best possible conditions for a phase-pure material. These data provide a 'fingerprint' of the compound because the $d$-spacings are fixed by the geometry of the crystal and the intensities are dependent on the contents of the unit cell. Hence, $d$–$I$ data may be used for identification of unknown materials by locating matching $d$–$I$ data in the PDF with the $d$–$I$ pairs obtained from an unknown specimen. Identification is the most common use of the PDF, but the presence of considerable supporting information for each entry in the file allows further characterization of the specimen. Examination of the crystal data, Miller indices, intensity values, scale factors, physical property data and the

comprehensive literature reference data provide extraordinarily useful information concerning the specimen under study.

To meet the needs of the scientific community, the PDF is continually revised and updated with new and improved information. Both improved instrumentation and more sophisticated analyses have led to a need for greater accuracy for the reference information. The task is exceedingly difficult because uncertainties in the experimental reference information and uncertainties in the unknown specimen need to be considered. It is a fitting tribute to the volunteer members of the ICDD that editorial activities have led to a comprehensive collection of powder diffraction data for which the quality of the entries is unmatched by other collections.

In 1978 the ICDD was formed as a not-for-profit corporation and reflected the vision of the founding members that this member-oriented organization should be truly international. The JCPDS held 30 members, whereas the current membership is 321, with the majority of members from outside the United States. The ICDD also began to support development activities, including programs for the improvement of the quality of the diffraction data. This effort is implemented, in part, by sponsoring a Grants-in-Aid (GiA) program. The GiA program is a competitive financial assistance program designed to encourage scientists working on new phases to submit high-quality diffraction data for inclusion in the PDF, and for the production of new patterns of phases of current interest, or preparation of the phases themselves. As a result of these grants, the ICDD receives a continual flux of new and potentially technologically relevant entries into the PDF. The ICDD currently supports 57 GiA research groups in seven countries.

In 1978 the production of 'cards' to depict entries was discontinued, with books and microfiche (although the latter has since been discontinued) as the only hardcopy forms published. The PDF-2 database was published on CD-ROM in 1987, reflecting the pragmatic need for storage media with large capacity mixed with visionary expectations for this new media. CD-ROMs were produced before CD-ROM reader standards had been universally adopted. Early distributions of the PDF-2 required readers with proprietary interfaces. This meant that users of the PDF-2 purchased the associated CD-ROM drives with which to read the data.



46-1045 ★

$SiO_2$

Silicon Oxide

Quartz, syn

Rad. CuKα1 λ 1.540598 Filter Ge Mono. d-sp Diff.
Cut off Int. Diffractometer $I/I_{cor}$ 3.41
Ref. Kern, A., Eysel, W., Mineralogisch-Petrograph. Inst., Univ. Heidelberg, Germany, *ICDD Grant-in-Aid*, (1993)

Sys. Hexagonal S.G. $P3_221$ (154)
a 4.91344(4) b c 5.40524(8) A C 1.1001
α β γ Z 3 mp
Ref. Ibid.

$D_x$ 2.65 $D_m$ 2.66 SS/FOM $F_{30}$=539(.002,31)
ωε nωβ 1.544 εγ 1.553 Sign + 2V
Ref. Swanson, Fuyat, *Natl. Bur. Stand. (U.S.), Circ. 539*, **3** 24 (1954)

Color White
Integrated intensities. Pattern taken at 23(1) C. Low temperature quartz. 2θ determination based on profile fit method. $O_2Si$ type. Quartz group. Silicon used as internal standard. PSC: hP9. To replace 33-1161. Structure reference: *Z. Kristallogr.*, **198** 177 (1992).

See following card.

| dÅ | Int | hkℓ | dÅ | Int | hkℓ |
|---|---|---|---|---|---|
| **4.2550** | 16 | 100 | 1.1530 | 1 | 311 |
| **3.3435** | 100 | 101 | 1.1407 | <1 | 204 |
| 2.4569 | 9 | 110 | 1.1145 | <1 | 303 |
| 2.2815 | 8 | 102 | 1.0816 | 2 | 312 |
| 2.2361 | 4 | 111 | 1.0638 | <1 | 400 |
| 2.1277 | 6 | 200 | 1.0477 | 1 | 105 |
| 1.9799 | 4 | 201 | 1.0438 | <1 | 401 |
| **1.8180** | 13 | 112 | 1.0346 | 1 | 214 |
| 1.8017 | <1 | 003 | 1.0149 | 1 | 223 |
| 1.6717 | 4 | 202 | 0.9896 | <1 | 115 |
| 1.6592 | 2 | 103 | 0.9872 | <1 | 313 |
| 1.6083 | <1 | 210 | 0.9783 | <1 | 304 |
| 1.5415 | 9 | 211 | 0.9762 | <1 | 320 |
| 1.4529 | 2 | 113 | 0.9608 | <1 | 321 |
| 1.4184 | <1 | 300 | 0.9285 | <1 | 410 |
| 1.3821 | 6 | 212 | 0.9182 | <1 | 322 |
| 1.3750 | 7 | 203 | 0.9161 | 2 | 403 |
| 1.3719 | 5 | 301 | 0.9152 | 2 | 411 |
| 1.2879 | 2 | 104 | 0.9089 | <1 | 224 |
| 1.2559 | 3 | 302 | 0.9009 | <1 | 006 |
| 1.2283 | 1 | 220 | 0.8972 | <1 | 215 |
| 1.1998 | 2 | 213 | 0.8889 | 1 | 314 |
| 1.1978 | <1 | 221 | 0.8814 | <1 | 106 |
| 1.1840 | 2 | 114 | 0.8782 | <1 | 412 |
| 1.1802 | 2 | 310 | 0.8598 | <1 | 305 |

**Figure 1**
Example of a PDF-2 image for quartz. The star in the upper right corner indicates a pattern of high quality.

There are a number of databases available to the X-ray diffraction community. Two of the more complete of these (for inorganic compounds) are the PDF, maintained by the International Centre for Diffraction Data (ICDD), and the Inorganic Crystal Structure Database (ICSD; Belsky *et al.*, 2002), maintained by Fachinformationszentrum (FIZ). While these databases have proven their usefulness in a wide range of applications, there has been little attempt to exploit a combination of these databases. Following an initiative on the part of the ICDD, an agreement has been made between the ICDD and FIZ of Karlsruhe, Germany, that allows mutual use of the PDF and the ICSD databases. Three major advantages have accrued from this cooperation. First, by use of cross-reference 'hooks' for each database entry, the user has access to experimental powder data from the PDF and structural information from the ICSD, permitting the full modeling of the experimental pattern. Secondly, the PDF can be supplemented using powder patterns calculated from structural data in the ICSD. Thirdly, the combined efforts of the two different editorial groups can only help to improve the overall content and quality of diffraction data. This paper considers the development and use of multiple databases.

## 2. The ICDD Powder Diffraction File

The PDF is a collection of single-phase X-ray powder diffraction patterns in the form of tables of the interplanar spacings ($d$) and relative peak intensities $I$(rel) characteristic of the compound. The PDF has been used for almost five decades (Hanawalt, 1983) and the ICDD maintains the PDF by continually adding new and updated diffraction patterns to the collection. In the early 1980s, the ICDD's editorial system was automated to allow detailed reviews to be made of all new patterns entering the PDF. Also, to assure the quality of existing data in the PDF (data added to the file previous to this time), the ICDD initiated a critical review of all numerical data in the PDF for Sets 1–32 (Wong-Ng *et al.*, 1982). This initial review process has now been extended to all data (recent and historical). Currently 2500 such patterns are added each year, comprising approximately 1900 inorganic patterns and 600 organic patterns. There is a continuing effort by the ICDD to ensure that new patterns being added to the PDF contain a significant proportion of phases that represent current needs and trends in industry and research. The master database of powder patterns is continually undergoing revision and updating, but in order to ensure that all database users have the opportunity to work with the same version, a 'frozen' version of the master database is produced each year and is supplied as the PDF-2.

The PDF-2 contains a series of individual data sets. Fig. 1 shows the layout of a typical PDF-2 image. As will be seen, each individual data set in the file contains, as a minimum, a list of $d$–$I$ pairs, chemical formula, name, a unique identifica-

**Table 1**
Subfiles of the PDF.

Note that Sets 1–47 were issued in August 1997 and Release 2001 was issued in August 2001.

| Subfile | Entries in Sets 1–47 | Entries in Release 2001 |
|---|---|---|
| Inorganic | 47 797 | 114 546 |
| Organic | 19 399 | 24 133 |
| Metals and alloys | 12 750 | 28 737 |
| Minerals | 4404 | 15 817 |
| Cement | 381 | 404 |
| Common phases | 3200 | 3826 |
| Corrosion products | 14 440 | 60 520 |
| Detergents | 2 | 2 |
| Dyes and pigments | 251 | 330 |
| Educational | 1071 | 1071 |
| Explosives | 176 | 240 |
| Forensic materials | 3680 | 3757 |
| Pharmaceuticals | 153 | 1985 |
| Polymers | 399 | 610 |
| Superconductors | 1135 | 2691 |
| Zeolites | 898 | 1764 |
| Totals in Sets | 65 907 | 136 895 |

tion (PDF) number and a reference to the primary source. In addition to this information, supplemental data may be added where available, including: Miller indices for all lines, unit-cell and space-group data, physical constants, experimental details and other comments. Because the number of patterns in the PDF is large, special ways of organizing the *d* and *I* data into subfiles have been devised; the more important of these subfiles are listed in Table 1. A smaller version of the database, called PDF-1, was produced mainly for search programs on minicomputers with limited disk storage. Up to Set 47, the 65 907 patterns of the PDF-2 database require *ca* 186 Mbytes of storage, and PDF-1 *ca* 39 Mbytes.

## 3. Other crystallographic databases

As shown in Table 2, there are a number of databases available which record the results of X-ray diffraction work. The majority of these databases are designed and maintained for the single-crystal community rather than for the powder community. Nevertheless, much cross fertilization can and does take place. For example, a number of the patterns in the PDF are calculated from single-crystal data of the type contained in the databases listed in Table 2. As we shall see later, a mutual agreement between ICDD and CCDC will lead to a completely new PDF for organic materials.

## 4. Phase identification by X-ray powder diffraction

Since every crystalline material gives, at least in principle, a unique X-ray diffraction pattern, the study of diffraction patterns from unknown phases offers a powerful means of qualitative identification, by comparing an X-ray pattern from the material to be analyzed with a file of single-phase reference patterns (see *e.g.* Jenkins & Snyder, 1996). Although the potential for qualitative phase identification was certainly recognized in the very early days of X-ray diffraction, the first attempts to list reference patterns were not published in detail until the late 1930s (Hanawalt *et al.*, 1938), along with means of archiving and retrieval of patterns. These methods still provide the basis of many search/match methods in use today. All databases require some type of index system to allow access to information contained within the database (even paper products were distributed with a combination of Index plus Search Manual for each of the main subsets of the PDF). A number of manual searching methods have been developed over the last 40 years, based on the three methods in common use today: the Alphabetic method, the Hanawalt method and the Fink method.

The Alphabetical Index is designed to permit a rapid systematic search for all patterns with a specified chemical content. The Hanawalt method involves grouping the patterns in the PDF according to the *d* value of the 100% intensity line. Each interval is often sorted on the *d* value of the second most intense line. Subsequent lines are listed in order of decreasing intensity. The Fink system indexes a pattern on its eight largest *d*-spacing lines and eight separate entries are made using a rotation of *d* values. Each year an Alphabetical Index and a Hanawalt Search Manual are published. A Fink Search Manual is published at non-regular intervals. Table 3 illustrates the various entry methods employed in the common indexes. It will be seen that the Alphabetic Index is a chemistry-based index using only elemental information. The Hanawalt Index is an intensity driven index since it employs only the strongest lines for searching. The Fink Index, on the other hand, is a *d*-spacing driven index since it employs mainly the largest *d* values. There have been numerous attempts to automate the search/matching process completely, resulting in several very successful commercial products (see Jenkins *et al.*, 1979; Snyder *et al.*, 1979; Snyder, 1982), which now permit routine external and internal standard calibration, precision alignment checks, and new levels of accuracy and precision in data collection and analysis. The ICDD has recently introduced a search-index program, *PCSIWIN* (Faber *et al.*, 2001), that computerizes Alphabetic, Hanawalt and Fink searches as described above.

The ability to recognize a reference pattern in an unknown pattern strongly depends on the quality of the *d* and *I* data of both the reference material and the unknown sample. One of the principal problems in the identification of materials by comparison of an experimental pattern with reference patterns is the variability in the quality of the data. The experimental technique used to measure the pattern is one of the first quality indications to a user of a reference database. For Debye–Scherrer camera data, one should assume an error window of $\Delta 2\theta = 0.1^\circ$, for normal diffractometer data one typically assumes a $\Delta 2\theta = 0.05^\circ$ window, and for internal-standard-corrected diffractometer or Guinier camera data a $\Delta 2\theta$ window as low as $0.01^\circ$ may be assumed. The other point to be considered is that although the experimentally measured parameter is generally the $2\theta$ value, the search/match parameter is invariably the *d* value. Unfortunately, the error relationship between $2\theta$ and *d* is nonlinear, and the most

**Table 2**
Crystallographic databases.

| Name | Content and reference | Center |
|---|---|---|
| Cambridge Structural Database (CSD) | Organic and metallo-organic structures (Allen, 2002) | CCDC, Cambridge, UK |
| Inorganic Crystal Structure Database (ICSD) | Inorganic structures (Belsky *et al.*, 2002) | FIZ, Karlsruhe, Germany |
| Metals Data File (CRYSTMET) | Metals, intermetallics and alloy structures (White *et al.*, 2002) | Toth Information Systems, Ottawa, Canada |
| Nucleic Acid Database (NDB) | Nucleic acid structures (Berman *et al.*, 2002) | Rutgers University, USA |
| Protein Data Bank (PDB) | Macromolecular structures (Berman, 2002) | Research Collaboratory for Structural Bioinformatics (RCSB), Rutgers University, USA |
| NIST Crystal Data (CD) | Inorganic and organic unit cell data | NIST, Gaithersburg, USA |
| Pauling File | Non-organic crystalline-state materials (currently metals and binary alloys) | JST, MPDS and RACE |

useful lines for phase identification are the low-angle lines, which are also the lines subject to the largest error in $d$.

As the quality of both reference and experimental patterns improves, the problem of pattern recognition becomes easier. There has been a major effort over the past several years to meet the ever-increasing demand for the higher quality data needed as a result of improved instrumentation and better techniques (Jenkins *et al.*, 1987). The introduction of the computer for data collection, treatment and processing has improved the quality of measured $d$-spacings, leading to an ongoing need for improvement in the quality of reference patterns. As an example, the modern automated powder diffractometer offers the user the possibility of producing $d$-spacing accuracies of *ca* 1 part per 1000 for all but the larger $d$ values. This quality of data corresponds to an average angular error in $2\theta$ of 0.01 to 0.03°. Whether the user obtains this quality still depends on the sample preparation and calibration; however, the quality of data in the PDF has improved with time.

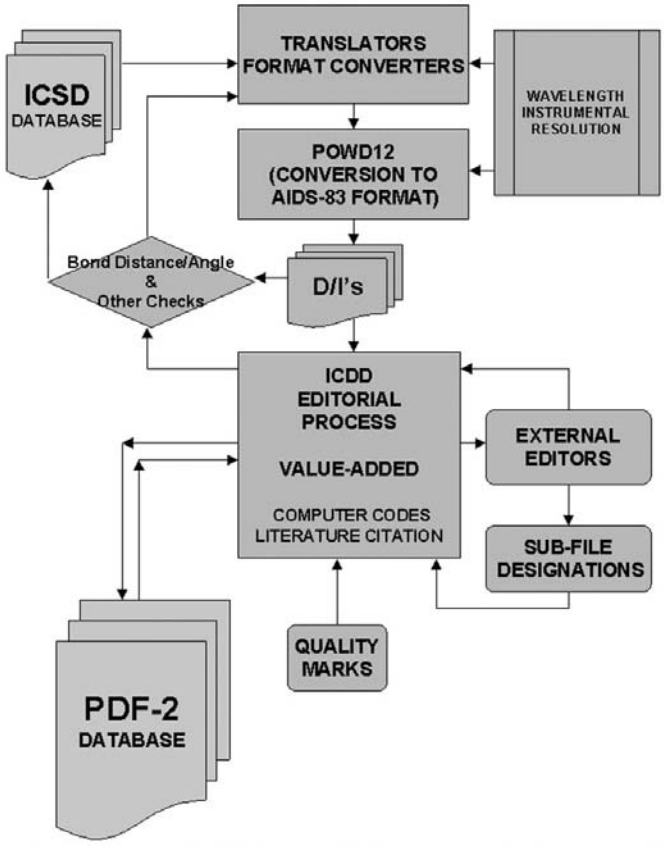



**Figure 2**
Editorial process employed for the production of calculated X-ray patterns.

Two methods are employed in the PDF to quantify experimental data quality. The first of these is a data quality mark that appears at the top right-hand corner of the card image. A 'Star' quality pattern indicates that the data are of the best quality, with an average absolute $\Delta 2\theta$ value of $\leq 0.03°$, all lines in the pattern having been indexed and the intensities measured quantitatively. An 'I' quality pattern has been indexed with no more than two lines being unaccounted for, with an average absolute $\Delta 2\theta$ value of $\leq 0.06°$ and again the intensities having been measured quantitatively. A 'C' quality pattern means that the $d$ and $I$ values have been calculated from atomic parameters. In this case, $2\theta$ is not defined. We should bear in mind, however, that calculated patterns depend upon the lattice parameters that are obtained from experiment. This means that uncertainty between reference patterns and those obtained from experiment will still be seen. An 'O' pattern is of low precision, poorly characterized and often has no unit-cell data. The possibility also exists that an 'O' pattern might contain a mixture. A 'Blank' generally indicates a pattern that does not meet the criteria for a 'Star', an 'I' or a 'C'. When no unit-cell data for such a pattern are available, it is impossible to assess the accuracy of individual lines in the pattern. The 'R' quality mark is used for patterns where it is clear that the $d$ values are directly the result of Rietveld refinement of the data. When unit-cell data are available, permitting its computation, the $F(N)$ figure of merit is also included in the database (Smith & Snyder, 1979).

## 5. Inclusion of calculated powder patterns in the PDF

The calculation of powder patterns from crystallographic structural data has reached some degree of maturity and a number of software programs are available to perform such calculations (*e.g.* Clark *et al.*, 1973; Yvon *et al.*, 1977). It has been the ongoing practice of the ICDD to include calculated patterns in the PDF where this was felt to be appropriate. As of Set 47, there were about 3700 calculated patterns in the

**Table 3**
Types of PDF data searching indices.

| Index | Entry method | Search parameters |
|---|---|---|
| Alphabetic | Chemistry | Permuted elemental symbols |
| Hanawalt | *I*/*d* | Three strongest lines |
| Fink | *d*/*I* | First eight longest lines |
| EISI (elemental and interplanar spacing index) | Chemistry/*d* | Low-/high-*Z* elements; *d*-spacing, used in electron diffraction since *I* values are absent |
| Boolean | Various | *d*-spacings, chemistry, strong lines, CODEN, physical properties, functional groups *etc.* |

**Table 4**
Statistics on the calculation of powder patterns from the ICSD database as given in Release 2001.

| Records | Sets 1–47 | Release 1998 | Release 2001 |
|---|---|---|---|
| Total number of entries | 65 907 | 125 342 | 136 896 |
| Number of calculated patterns from ICSD | 0 | 37 831 | 49 384 |
| Number of entries with *I*/*I*(cor) | 5105 | 45 232 | 56 781 |

PDF. The reasons for the inclusion of calculated patterns include the following:

(i) The availability of both experimental and calculated patterns allows evaluation of experimental preferred-orientation problems (McCarthy *et al.*, 1992).

(ii) Experimental and specimen-related effects, such as instrumental resolution and size, strain and defect distortions of the line shapes, may, at times, obscure details in the experimental pattern.

(iii) Experimental data may not be available (Cantrell *et al.*, 1988).

(iv) Simulation of ill-ordered materials is possible (Reynolds, 1989).

(v) To combine data so as to supplement incomplete PDF entries.

The 1998 release of the PDF was enhanced by the addition of approximately 40000 patterns obtained from the ICSD database by calculation. This enhancement does not require that users have an ICSD license: the calculated patterns are a permanent addition to the PDF. All of the new calculated patterns will have passed through the preliminary editorial process. However, they will not have passed through the complete editorial process as would happen with all calculated patterns in the PDF before Set 48. In order to identify those patterns that have only partially completed the full editorial review, they are given either a 'MAP' (modeled additional pattern) or a 'MIP' (modeled in process) subfile mark. There is an ongoing, but time-consuming project to pass all calculated patterns through the full editorial review process, after which stage a 'MAP' pattern will have a subfile code assigned. An overall scheme of the editorial process for calculated patterns is given in Fig. 2.

The very existence of the ICDD is the result of the diffraction community's requirement for experimental patterns for phase identification. Because of the many specimen and instrumental related distortions of a powder pattern, phase identification will often fail without an edited, high-quality experimental database showing patterns as they have actually been observed. The existence of the comprehensive collection of inorganic crystal structure information contained in the ICSD permits the creation of a supplemental set of calculated patterns that can aid and extend the ability to identify materials and extract the full amount of information contained in the diffraction pattern of a specimen. In addition to the potential enhancements to phase identification strategies, there is an immediate benefit of adding about 20000 calculated patterns for phases that are not represented by experimental patterns in the PDF. While it is true that these new phases may differ significantly from experimental patterns, the presence of these calculated patterns clearly enhances the PDF.

Traditionally, the ICDD has employed a logical numbering process for giving each PDF pattern a unique number. This is done by subtracting 1950 from the year of a given annual update then adding a further 4 digits from 0001 to 2500. For example, the first pattern in the 1997 update is 47–0001 and the 2500th is 47–2500. In order to retain this system, the ICDD has

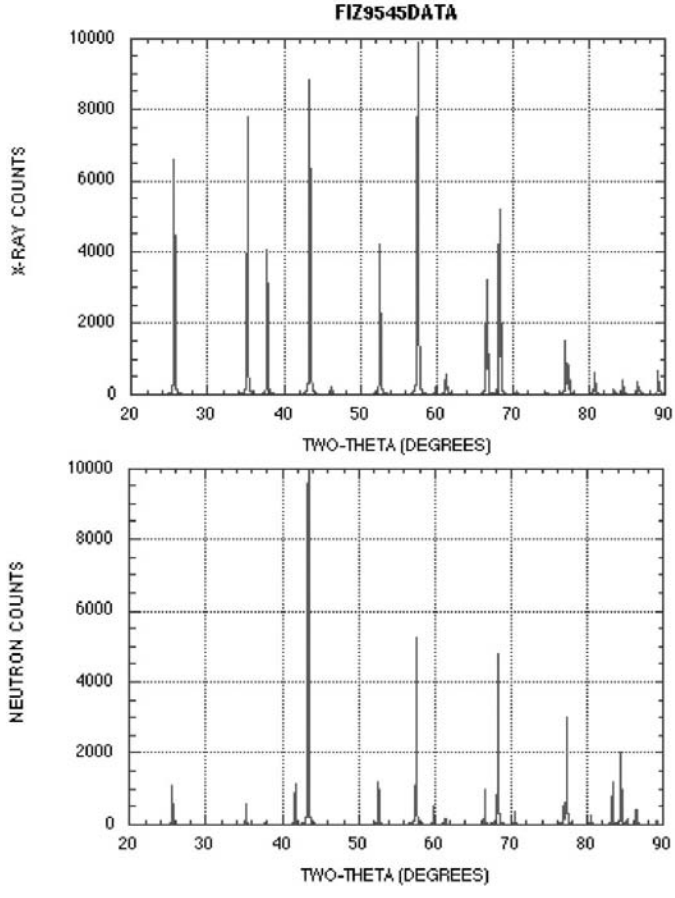



**Figure 3**
Example of powder patterns for corundum, calculated as X-ray data and neutron data.

**Table 5**
Comparison of PDF-2 and PDF-4 database structures.

| | PDF-2 | PDF-4 |
|---|---|---|
| File format | ASC-II, 80 character records | Binary, encrypted |
| Entry format | Variable-length record sets | Fields and entries in tables |
| Entry expansion | Additional record sets | Additional entries |
| Accommodation of new information | Limited by format of record set, fixed | Additional fields and tables easily added |
| Indexing | Either bit-mapped or vector-mapped indexes for searchable quantities | Indexing is attached to the table structures |
| Indexing location | Built and located outside the database | Automatically rebuilt as table structures evolve |
| Security | Inherently none | Many security measures can be implemented |

allocated all calculated 'MAP' patterns a unique number of the form 70–*xxxx*, in which *xxxx* will vary from 0001 to 2500. When 70–2500 is reached, the next pattern is given the number 71–0001 *etc*.

One of the advantages to be gained by the use of calculated patterns is that it is, in principle, not difficult to calculate a powder pattern as if it had been recorded with different sources. These include the digitized trace PDF-3, neutron diffraction, electron diffraction, plus any wavelength for X-ray data. As an example, Fig. 3 shows powder patterns calculated for corundum, assuming first X-ray data and second neutron data. As can be seen, there are enormous differences in the scattering contrast for X-rays and neutrons. In the neutron case, the O atoms are the 'heavy-atom scatterer'. In addition, one can calculate derived data, such as the $I/I$(cor) value (Hubbard *et al.*, 1976). As shown in Table 4, as of Release 2001, there are about 57000 data sets with $I/I$(cor) data included.

The enhanced database will follow the same format as the previous PDF-2 database. Table 4 shows what the combination database contains. It is anticipated that this product will be distributed, in the short term, using conventional CD-ROM technology. However, the maximum capacity of the CD-ROM is fast being approached and we are now exploring the feasibility of alternative distribution media, particularly DVD technology.

## 6. The role of fully digitized powder patterns (PDF-3)

The ICDD has been archiving fully digitized raw data sets, PDF-3, for a number of years. The availability of the trace of the original experimental data can often give useful background information for the editorial process. In addition, a new strategy has recently been introduced (Smith *et al.*, 1988; Caussin *et al.*, 1988) which has dramatically improved the success rate of the search/match process. This strategy is based on searching the whole observed pattern with its background, not just the *d–I* list, and on adding candidate phases together to compose, rather than decompose, an observed multi-phase pattern.

While the availability of such raw data is clearly useful, there are a number of problems in attempting to provide a user version that is reasonably accurate and reasonably complete. The main problem in ensuring the accuracy and precision of raw data is the proliferation of instrument types in use today. Since one would wish to compare their data set with someone else's data set, an independent means of calibration is required that would allow reduction of all data sets to a common base. Following extensive round robin tests (Schreiner *et al.*, 1992; Valvoda *et al.*, 1995) to quantify the problem, the ICDD has developed a procedure whereby all individual data sets are related to a data set taken on NIST SRM 1976 (Cline *et al.*, 1992). The second major problem is in compiling a database that is reasonably complete. The PDF-2 started in the mid 1940s and it has taken more than 50 years to accumulate the current number of phases. Until very recently, virtually none of these reduced patterns had archived original scans. Up to this point, the ICDD has concentrated its efforts to accumulate fully digitized patterns for those compounds where line shapes are invaluable in the identification process, *e.g.* clays, polymers *etc.* With the potential to produce full patterns *via* the pattern calculation process, the number of potential PDF-3 patterns increases dramatically.

## 7. Database structures

The traditional method for the storage of data is to reduce the experimental pattern to a table of *d–I* values, often referred to as a reduced pattern because the process of data treatment reduces the large volume of data in the raw scan to a concise digital form. Unfortunately, during the data reduction process, much information concerning the line shape and intensity distribution is lost. Although it may be more useful in some cases to utilize the full diffractogram, until recently storage limitations have inhibited the development of a pattern reference file of fully digitized patterns. Whether the pattern is a complete digitized pattern or a reduced pattern, in either case it is necessary to archive the data in a convenient form. The first effort to computerize the PDF dates back to the early 1970s and at that time the (then) National Bureau of Standards (NBS) developed a flexible and versatile data editing, checking and archiving program, which in its latest execution is called *NBS*AIDS*83 (Mighell *et al.*, 1981). Over the past 20 years, the *NBS*AIDS*83 program has become the backbone of the ICDD in-house editorial process. More recently, an initiative on the part of the IUCr has resulted in the acceptance of a general-purpose data transfer format, the CIF (Hall *et al.*, 1991). This format was based on the Self-Defining Text Archive and Retrieval (STAR) method suggested by Hall (1991). Programs have now been developed by the ICDD for

**Table 6**
Current production activity for PDF-4.

| Product designation | Number of entries | Release date |
|---|---|---|
| PDF-4/Metals and Alloys RDB | 24 096 | October 2000 |
| PDF-4/Minerals RDB | 14 600 | August 2001 |
| PDF-4/Metals And Alloys 2001 RDB | 28 700 | October 2001 |
| PDF-4/Full File 2002 RDB | 136 000 | Spring 2002 |
| PDF-4/Organic 2002 RDB | 150 000 | Late fall, 2002 |

transfer from AIDS to CIF format and *vice versa*. In the 1980s there was a similar initiative on the part of the molecular spectroscopy community to provide an archival system for fully digitized spectra. This system was called JCAMP (McDonald & Wilks, 1988) and, in a slightly revised form, was used for a number of years by the ICDD for the archiving of PDF-3 data. More recently, the PDF-3 data have been converted to CIF-STAR. Thus, over the past several years, the CIF-STAR format has become the archival database format of choice for the ICDD.

## 8. New relational database (RDB) structures (PDF-4)

The ICDD is transitioning to relational database (RDB) structures to house the PDF; these are designated PDF-4. Perhaps the best way to understand the need for RDBs is to discuss the limitations of the flat-file structure defined by the PDF-2 (using the *NBS*AIDS83* format). In Table 5 we show a comparison of database features for PDF-2 and PDF-4. From Table 5 we can see several important advantages for PDF-4 RDBs. The RDBs are easily extensible as new data records are appended to the database. The index structures are located within the tables resident in the RDB. As such, the index files are rebuilt on-the-fly. This means that as records are edited or appended, the index files are automatically generated. New properties can be easily appended in the RDB since this operation corresponds to appending new fields to existing or new tables. RDBs are typically housed within commercial software containers and this means that dependencies between software houses and database organizations are increased. However, perhaps the most exciting development for PDF-4 RDBs is that this represents a first step towards an ability to perform total pattern analyses and continues the technical evolution of the PDF from handwritten files to searchable databases.

To illustrate this more clearly, Fig. 4 shows a fully digitized diffraction pattern synthesized on-the-fly from the PDF-4/Organic RDB. The PDF entry is defined as 57–2056 and the corresponding structural data can be found in the Cambridge Structural Database (CSD), using reference code CUASPR. Although not illustrated in Fig. 4, settings options are present to account for particle size and strain effects. The extension of these ideas to account for preferred orientation and texture is straightforward. At this stage, the PDF-4 supplies peak intensities and *d*-spacings as a starting point, and also displays the total pattern. All of this taken together [along with $I/I$(cor) discussed above] sets the stage for total pattern analysis including, among others: phase identification; quantitative phase analysis; anisotropic size and strain information; preferred orientation; crystallinity.

The ICDD and the Cambridge Crystallographic Data Centre (CCDC) have reached an agreement to produce a new organic database, designated PDF-4/Organic 2002. Scheduled for release in November 2002, this database will contain 150000 entries, 25000 experimental patterns, and 125000 calculated patterns from the CSD, with cross references between the calculated patterns in the PDF-4/Organic database and the CSD. This database will provide $I/I$(cor) and fully digitized powder patterns for approximately 130000 entries. Display of two-dimensional structures for almost all entries is also provided.

To summarize this transition from PDF-2 to PDF-4 RDBs, the current production activity for PDF-4 is illustrated in Table 6.

## 9. Use of multiple databases

A major benefit to be gained by the simultaneous use of multiple databases, and the one which potentially has the

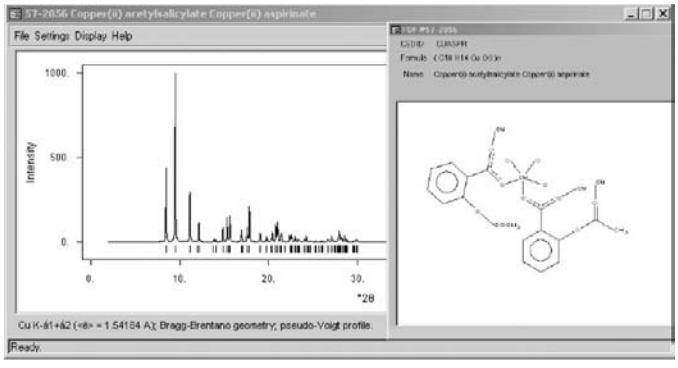



**Figure 4**
Example of an on-the-fly calculated powder pattern for copper(II) acetylsalicylate copper(II) aspirinate. An inset shows the two-dimensional structure of this entry in the PDF-4/Organic RDB. The cross-referenced CSD reference code for the entry is CUASPR.

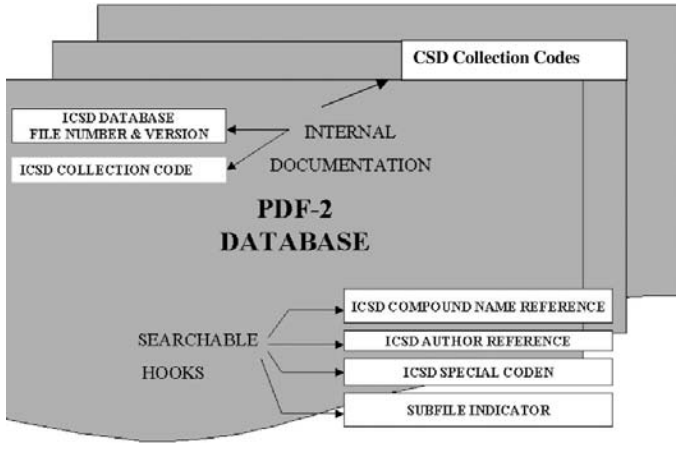



**Figure 5**
Example of use of pointers between different databases.

longest-term impact, comes from the cross-correlation of the PDF and the ICSD; see Fig. 5. Today, automated search/match algorithms are limited to listing the best-matched phases in the order of 'goodness of fit'. The automated ability to access the atomic coordinates and then generate the calculated patterns for potential phases identified in an unknown mixture opens a new era in phase identification. Least-squares refinement of the calculated patterns will permit the next generation of algorithms to test and resolve postulates concerning preferred orientation and solid-solution shifting in establishing the match. With this new ability, algorithms will be able to identify, fully automatically and unambiguously, the actual phases in an unknown, when the appropriate information is in both of the databases. In addition, all of the other information potentially contained in the powder patterns can be routinely extracted as part of the phase identification: semi-quantitative analysis from the calculated $I/I$(cor) values, concentration of components in identified solid solutions, all degrees of preferred orientation in a specimen, the crystallite size and strain of each of the phases exhibiting line broadening *etc*. The integration of the crystal structure information with the PDF will bring about a new era of phase analysis for licensed users of both databases.

In a companion paper (Kabekkodu *et al.*, 2002), we have assembled several cases involving data mining using our RDB databases that emphasize the flexibility and power of using RDB structures for the PDF.

Editing, producing and calculating powder patterns requires significant team effort by the ICDD employees, volunteer ICDD members and the ICDD Board of Directors. We thank the hundreds of researchers and scientists who have contributed to the organization over the past 60 years.

## References


Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.

Belsky, A., Hellenbrandt, M., Karen, V. L. & Luksch, P. (2002). *Acta Cryst.* B**58**, 364–369.

Berman, H. (2002). *Acta Cryst.* B**58**, 899–907.

Berman, H., Westbrook, J., Feng, Z., Iype, L., Schneider, B. & Zardecki, C. (2002). *Acta Cryst.* B**58**, 889–898.

Cantrell, J. S., Beiter, T. A. & Sullenger, D. B. (1988). *Adv. X-ray Anal.* **31**, 371–376.

Caussin, P., Nusianovici, J. & Beard, D. W. (1988). *Adv. X-ray Anal.* **31**, 423–430.

Clark, C., Smith, D. K. & Johnson, G. G. Jr (1973). *A Fortran IV Program for Calculating X-ray Powder Patterns*. Department of Geosciences, The Pennsylvania State University, State College, PA, USA.

Cline, J., Schiller, S. B. & Jenkins, R. (1992). *Adv. X-ray Anal. A*, **35**, 341–352.

Faber, J., Weth, C. A. & Jenkins, R. (2001). *Mater. Sci. Forum*, **378–381**, 106–111.

Hall, S. R. (1991). *J. Chem. Inf. Comput. Sci.* **31**, 326–333.

Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *Acta Cryst.* A**47**, 655–685.

Hanawalt, J. D. (1983). *History of the Powder Diffraction File*, in *Crystallography in North America – Apparatus and Methods*, ch. 2, pp. 215–219. American Crystallographic Association.

Hanawalt, J. D., Rinn, H. W. & Frevel, L. (1938). *Ind. Eng. Chem. Anal. Ed.* **10**, 457–512.

Hubbard, C. R., Evans, E. H. & Smith, D. K. (1976). *J. Appl. Cryst.* **9**, 169–174.

Jenkins, R., Hahm, Y., Pearlman, S. & Schreiner, W. N. (1979). *Norelco Rep.* **26**, 1–15.

Jenkins, R., Holomany, M. & Wong-Ng, W. (1987). *Powder Diffr.* **2**, 84–87.

Jenkins, R. & Snyder, R. L. (1996). *X-ray Powder Diffractometry*. New York: John Wiley.

Kabekkodu, S. N., Faber, J. & Fawcett, T. (2002). *Acta Cryst.* B**58**, 333–337.

McCarthy, G. M., Martin, K. J., Holzer, J. M. & Grier, D. G. (1992). *Adv. X-ray Anal.* **35**, 17–23.

McDonald, R. S. & Wilks, P. A. (1988). *Appl. Spectrosc.* **42**, 151–162.

Mighell, A. D., Hubbard, C. R. & Stalick, J. K. (1981). *Natl*AIDS*80: *A Fortran Program for Crystallographic Data Evaluation*. Technical Note 1141. National Bureau of Standards, Washington, DC, USA.

Reynolds, R. C. Jr (1989). *Rev. Miner.* **20**, 145–181.

Schreiner, W. N., Jenkins, R. & Dismore, P. F. (1992). *Adv. X-ray Anal.* **35**, 333–340.

Smith, D. K. & Jenkins, R. (1996). *J. Res. Natl Inst. Stand. Technol.* **101**, 259–271.

Smith, D. K., Johnson, G. G. Jr & Wims, A. M. (1988). *Aust. J. Phys.* **41**, 311–321.

Smith, G. S. & Snyder, R. L. (1979). *J. Appl. Cryst.* **12**, 60–65.

Snyder, R. L. (1982). *Adv. X-ray Anal.* **24**, 83–90.

Snyder, R. L., Mallory, C. L., Smith, S. T., Osgood, B. C. & Howard, S. A. (1979). *The Rebirth of X-ray Powder Diffraction*. New York State College of Ceramics Report, 17, Vol. III, No. 5.

Valvoda, V., Rafaja, D. & Jenkins, R. (1995). *Adv. X-ray Anal.* **39**, 572–577.

White, P. S., Rodgers, J. R. & Le Page, Y. (2002). *Acta Cryst.* B**58**, 343–348

Wong-Ng, W., Holomany, M., McClune, F. & Hubbard, C. R. (1982). *Adv. X-ray Anal.* **26**, 87–88.

Yvon, K., Jeitschko, W. & Parthé, E. (1977). *J. Appl. Cryst.* **10**, 73–74.