materials (conventional superconductors, superconductor reaction products, superconductor-related and high-T_c superconductors), terpenes and thermoelectric materials. There is an educational package for classroom use, and the complete PDF is available for educational use on a time-limited basis. A primary purpose of the subfile system is to limit the size of the search universe by applying prior knowledge of the system being studied. This greatly reduces the number of false positives in a database that contains hundreds of thousands of materials. Field experts are consulted to guide the criteria for subfile selection, allowing novices to use the subfiles without being a subject expert.

3.7.2.1. *Sources and formats of the PDF*

The data incorporated into the Powder Diffraction File are acquired through contributions from individual scientists, corporate laboratories, literature surveys and a Grant-in-Aid programme. Approximately 200 leading scientific journals are searched manually for powder-diffraction data. Additional literature surveys covering patents, dissertations and the remaining open literature are performed using various online resources and search techniques.

Release 2019 (the current release as of this writing) contains more than 893 400 unique material data sets. The large size and comprehensive coverage of the PDF is achieved through the ICDD's historical sources of powder data (searches of the original literature, contributions and the Grant-in-Aid programme) as well as current and historic collaborations with crystallographic database organizations. Each PDF entry is assigned a unique identifying number of the format *ss-mmmnnnn*. The integer *ss* indicates the source of the data: 00, ICDD location/generation of powder data; 01, Inorganic Crystal Structure Database; 02, Cambridge Structural Database; 03, NIST (a short-term collaboration focused on metals and alloys); 04, Pearson's Crystal Data; 05, ICDD extraction of atomic coordinates from published sources (including incommensurate/modulated structures). Powder-diffraction data for sources 01 through 05 are computed from the crystal structures provided by these sources.

The Powder Diffraction File is designed and produced in several different formats in order to serve different groups of users. The PDF-2 database is designed for phase identification of inorganic materials; many common organic materials have also been added to this database. The PDF-4+ database is the most advanced database and is designed for both phase identification and quantitative analysis. This database has comprehensive coverage of inorganic materials and contains numerous additional features such as digitized (raw) patterns, molecular graphics and atomic coordinates to facilitate Rietveld refinements. The PDF-4+ database is also available as a portable full-function WEBPDF-4+ version. The PDF-4/Minerals database is a subset of the PDF-4+ database, and is the most comprehensive collection of mineral diffraction data. The PDF-4/Organics database is designed for phase identification of organic and coordination compounds. It contains data from ICDD sources (both experimental powder patterns and extraction of coordinates) as well as patterns calculated from CSD entries.

Advances in hardware, software and computing power have led to the collection of higher-quality powder data, and thus have necessitated higher-quality reference data to perform more complex multiphase analyses and total-pattern analyses. The PDF now includes tools that permit users to evaluate different types of data collected using different types of detectors and different sources, including X-rays, neutrons and electrons. The goal is to include ideal specimen patterns in the PDF, patterns that can be modified by the user to correspond to the current experiment. The user can select the wavelength type and various instrumental parameters to simulate the whole diffraction pattern. A crystallite size calculation was added in 2007 and an orientation function in 2011.

Since 2006, the ICDD has begun to include several types of less-crystalline materials in the database, materials for which too much information is lost when reducing the raw data to a list of *d*-spacings and intensities. These materials include clays and other layered materials, mixed-crystallinity polymers, amorphous materials and nanomaterials.

Nanomaterials often contain crystalline and amorphous fractions, and their powder patterns are difficult to generate from an ideal crystal structure. The ICDD has developed quality-evaluation methods for noncrystalline materials, and has established two additional quality marks: 'good' (G) and 'minimal acceptable' (M). These marks reflect the quality of the supporting data used to characterize the material. An amorphous material with a G quality mark has been characterized by independent analyses verifying the stated composition or thermogravimetric/differential scanning calorimetry analyses confirming the physical stability or the presence of a glass transition. A G quality mark indicates that the editors are satisfied that the pattern is representative of both the diffraction conditions and the stated chemistry and have confidence that the user can reproduce the pattern using similar conditions. The quality mark M indicates that the ICDD received some supporting documentation but it was insufficient for structural interpretation and classification of the material.

Great care needs to be taken in interpreting the patterns of mixtures of crystalline and amorphous phases, particularly in the definition and subtraction of the background. Significant work is under way to develop and adapt numerical techniques for processing full patterns of low-crystallinity materials.

3.7.2.2. *Quality marks in the PDF*

All data are critically reviewed and evaluated by the PDF editorial staff. Each pattern must pass through a four-tiered editorial review process before it can be included in the PDF. As technology evolves, the quality requirements for reference data also evolve. As a result, the information in the PDF is continuously reviewed and upgraded for accuracy and quality.

For many years, a quality mark has been assigned to each experimental PDF entry. A Star (*/S) pattern represents high-quality diffractometer or Guinier data. Several criteria must be satisfied for a pattern to be assigned a Star quality mark:

(i) The chemical composition must be well characterized.
(ii) The intensities must have been measured objectively; no visual estimation is allowed.
(iii) The pattern has a good range and an even spread of intensities.
(iv) The completeness of the pattern is sensible.
(v) The *d*-spacing of each reflection with $d \leq 2.500$ Å is given to at least three decimal places. The *d*-spacings of reflections with $d \leq 1.2000$ Å are given to at least four decimal places.
(vi) No serious systematic errors exist.
(vii) The $|\Delta 2\theta|$ value (*i.e.* the difference between the observed peak position and the position calculated from the unit

cell) of a qualifying reflection is $\leq 0.05^\circ$. In the case of multiply-indexed reflections, only the minimum absolute $\Delta 2\theta$ is considered.

(viii) The average $|\Delta 2\theta| \leq 0.03^\circ$ for qualifying reflections.

(ix) No unindexed, space-group-extinct or impurity reflections are present.

An Indexed (I) quality mark indicates that the pattern has been indexed; therefore, the material is almost certainly single-phase. There is a reasonable range and spread of intensities, and the completeness of the pattern is sensible. The d-spacings of reflections with $d \leq 2.000$ Å have at least three significant figures after the decimal point. No serious systematic errors exist. No qualifying reflection has $|\Delta 2\theta| \geq 0.20^\circ$ and the average $|\Delta 2\theta|$ is $\leq 0.06^\circ$. The maximum number of unindexed, space-group-extinct or impurity reflections is two, but none of these reflections are among the eight strongest lines.

A Blank (B) quality mark represents a mid-range quality. An O quality mark means that the data have been obtained from a poorly characterized material or that the data are known (or are suspected) to be of low precision and accuracy. Such patterns include those from multiphase mixtures or from a phase that is poorly characterized chemically. The O mark is commonly assigned to patterns for which no unit cell is reported, unless qualifying information indicates a single-phase material. Usually, the editor will have inserted a comment to explain why the O mark was assigned. For patterns with a unit cell, the following criteria are used to suggest the presence of two or more phases: the number of unindexed, space-group-extinct or impurity reflections is ≥ 3, or one of the three strongest peaks is unindexed.

Beginning with Release 2006, the quality-mark system was extended to patterns calculated from structural data supplied by ICDD partners. The focus of the quality mark is to determine the confidence level of the structural model used and its impact on the calculated pattern (especially for the purpose of phase identification). The major step involves several crystallographic and editorial checks by the ICDD, followed by extraction and flagging of the warnings/comments in the structural databases. The resulting calculated patterns are classified based on the significance and nature of the warnings. Any possible corrections that can be applied to resolve the errors are performed before publishing the calculated pattern.

The crystallographic checking rules are designed based on the expected quality of a contemporary crystal structure. An estimate of the missing electron density is made based on the difference between the reported composition and the structural composition. Transformations of nonstandard space groups are checked; the reported site multiplicities must match those generated by the symmetry operators. All of the eigenvalues of the anisotropic tensor matrix for each atomic displacement must be positive. All anisotropic tensor coefficients must be permitted by the site symmetry. Displacement coefficients should fall in the range $0.001 < U < 0.1$ Å^2. Isotropic displacement coefficients must be positive. Mixed displacement coefficients are converted to a standard type. The reported value of Z must be consistent with the sum of the site multiplicities. Lattice parameters are checked for missing decimal points, missing standard uncertainties and the magnitudes of the uncertainties. R factors close to the theoretical limits (0.83 for centrosymmetric structures and 0.59 for non-centrosymmetric structures) are signs of potential errors in the conversion to/from absolute/percentage values. Site occupancies cannot be greater than 1. Refining part of the structure as a group without locating the positions of the constituent atoms (for example, in C_{60}) will generate a warning. Possible typographical errors in element symbols are checked by comparing the chemical formula, atomic coordinate list and chemical name. When a measured density is available, the percentage difference between the measured and calculated density is determined.

Many warnings/comments from the collaborating databases are used in assignment of the quality mark. Editorial comments on unusually short or long bond lengths or questionable bond angles are considered; the comment needs to be very specific for structures exhibiting disorder or partial/mixed occupancies. A listing of other types of comments considered is contained in the PDF-4+ database help documentation. Entries are assigned a quality mark of * (no warning found during data evaluation), I (minor warning), B (significant warning found), O (major warning), P (the structure was assigned by the editor based on a prototype) or H (hypothetical) according to the criteria in Table 3.7.1.

3.7.2.3. *Features of the PDF*

Most users access the PDF through the software provided by their instrument manufacturer, but it is a powerful standalone database. The PDF is a large relational database consisting of many linked tables. The complete set of features can be accessed through the PDF front end supplied by the ICDD. It is possible to directly access a PDF entry by entering its PDF number. However, one can search for an entry or a class of entries through a series of search tabs. Queries from multiple tabs can be combined in a single search, or individual searches can be saved in a history and combined using Boolean operations. The results of such searches can be analysed as a group or can be used as subfiles for *SIeve*, the search/index phase-identification add-on for the PDF.

Selections on the main search screen permit selection by the source of data, quality mark, primary/alternate, ambient/non-ambient and subfile or subclass. The comprehensive nature of the PDF means that there are often many entries for an individual material. The ICDD editorial staff and volunteer task groups assign one experimental and one calculated entry (if present) as primary entries for each phase so that the user can avoid the duplication if desired. The other entries are designated as alternates. The subfiles and subclasses provide a convenient means for the user to limit the size of the search universe based on prior knowledge and result in faster searches and fewer false-positive matches.

Perhaps the most commonly used screen is the Periodic Table tab for chemistry searches. Individual elements, groups, periods and pre-defined selections (nonmetals, semimetals *etc.*) can be selected and combined in various ways. The 'and' operation requires that all selected elements be present in the entries in the selection set, but other elements can also be present. The 'or' operation requires at least one of the selected elements to be present. The 'only' operation requires that all of the selected elements, and only those elements, be present in the hit. The 'just' operation results in a hit list of entries that contain the selected elements in all combinations: elements, binaries, ternaries *etc.* The results of these four types of element searches can also be combined using Boolean operations. An alternative way of using periodic-table screening is through the labelling of each element with 'yes', 'no' or 'maybe' to indicate elements that are known to be present, absent or unsure in the specimen.

The Formula/Name tab facilitates searches on formula, empirical formula, structural formula and formula type ANX [as in the Inorganic Crystal Structure Database (ICSD)]. The formulae may be exact or contain individual elements or strings.