entry can be located without any ambiguity and the best structure for the problem at hand can be used to start the Rietveld refinement.

3.7.5. Pearson's Crystal Data (PCD/LPF) (with Pierre Villars and Karen Cenzual)

3.7.5.1. *General information*

The Pearson's Crystal Data database (PCD; Villars & Cenzual, 2013) is an outgrowth of the (Linus) Pauling File (LPF; Villars *et al.*, 1998; http://www.paulingfile.com), which was designed to combine crystal structures, phase diagrams and physical properties under the same computer framework to form a tool useful for materials design. PCD is the result of a collaboration between Material Phases Data Systems (Vitznau, Switzerland) and ASM International (Materials Park, Ohio, USA). The retrieval software was developed by Crystal Impact (Bonn, Germany). As suggested by the name, Pearson's Crystal Data is a follow-up product to *Pearson's Handbook: Crystallographic Data for Intermetallic Phases* (Villars & Calvert, 1985, 1991; Villars, 1997). However, in contrast to the latter, it also covers oxides and halides, which represent about 80% of the compounds with more than four chemical elements.

The 2016/2017 release of Pearson's Crystal Data contains more than 288 000 data sets for more than 165 300 different chemical formulae, representing over 53 000 distinct chemical systems. To achieve this, the editors have processed over 93 500 original publications; recent literature is surveyed in a cover-to-cover approach, including about 250 journal titles. Over 153 000 database entries contain refined atom coordinates, as well as isotropic and/or anisotropic displacement parameters when published, whereas more than 72 000 data sets contain atom coordinates corresponding to the structure prototype assigned by the authors of the original publication or by the database editors. Approximately 15 000 data sets contain only crystallographic data such as the lattice parameters and possibly a space group.

When available in the original publications, each data set contains comprehensive information on the sample-preparation and experimental procedure, as well as on the stability of the phase with respect to temperature, pressure and composition. The presence of plots (cell parameters or diffraction patterns) in the original paper is indicated, and over 30 000 descriptions of the variation of the cell parameters as a function of temperature, pressure or composition are proposed. Roughly 18 300 experimental diffraction patterns are reported.

The Linus Pauling File was designed as a phase-oriented, fully relational database system. This required the creation of a 'distinct phases' table, with internal links between the three parts of the database. In practice, this means that the senior editors have evaluated the distinct phases existing in the system for every chemical system using all information available in the LPF. Each structure entry in Pearson's Crystal Data has been linked to such a distinct phase, which allows a rapid overview of a particular chemical system.

3.7.5.2. *Evaluation procedure*

Extensive efforts have been made to ensure the quality and reliability of the crystallographic data. Pearson's Crystal Data is checked for consistency by professional crystallographers, assisted by an original software package, *ESDD* (*Evaluation, Standardization and Derived Data*), containing more than 60 different modules (Cenzual *et al.*, 2000). The checking is carried out progressively, level by level. The following checks are made.

Individual database fields:

(i) order of magnitude of numerical values;
(ii) Hermann–Mauguin symbols, Pearson symbols;
(iii) consistency of journal CODEN, year, volume, first page, last page;
(iv) formatting of chemical formulae;
(v) neutrality of oxides and halides;
(vi) spelling.

Consistency within individual data sets:

(i) atom coordinates, Wyckoff letters, site multiplicities;
(ii) chemical elements in different database fields;
(iii) computed, published values (cell volume, density, absorption coefficient, *d*-spacings);
(iv) Pearson symbol, space group, cell parameters;
(v) Bravais lattice, Miller indices;
(vi) site symmetry, anisotropic displacement parameters.

Particular crystal-structure checks:

(i) interatomic distances, sum of atomic radii;
(ii) geometry of functional groups;
(iii) search for overlooked symmetry elements;
(iv) composition from refinement, chemical formula.

Consistency within the database:

(i) comparison of cell-parameter ratios for isotypic entries;
(ii) comparison of atom coordinates for isotypic entries with refined coordinates;
(iii) comparison of densities;
(iv) thorough search for duplicates, also considering translated references.

Wherever possible, misprints have been corrected based on arguments explained in remarks; as a result, more than 13 000 crystallographic data sets are accompanied by at least one erratum. In other cases remarks drawing the attention to discrepancies or unexpected features have been added.

The *ESDD* software package also produces derived data such as the Niggli reduced cell, equivalent isotropic displacement parameters, density and formula weight.

3.7.5.3. *Standardized crystallographic data*

The crystallographic data in Pearson's Crystal Data are presented as published, respecting the original site labels, but are also standardized following the method proposed by Parthé and Gelato (Parthé & Gelato, 1984, 1985; Parthé *et al.*, 1993). This second presentation of the same data has been further adjusted so that compounds crystallizing with the same prototype structure (isotypic compounds) can be easily compared. It is prepared in a three-step procedure as follows.

(i) The crystallographic data are checked for the presence of overlooked symmetry elements. Whenever it is possible to describe the structure in a higher-symmetry space group, or with a smaller unit cell, without any approximations, this is performed.
(ii) In the next step, the crystallographic data are standardized using the program *STRUCTURE TIDY* (Gelato & Parthé, 1987).
(iii) The resulting data are compared with the standardized data of the type-defining data set and, if relevant, adjusted using an *ESDD* module based on the program *COMPARE* (Berndt, 1994).

For data sets with no published coordinates, the cell parameters are standardized following the criteria defined for the unit-cell