

3. METHODOLOGY

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

3.7. CRYSTALLOGRAPHIC DATABASES

and space-group setting. For data sets with unknown space group, the cell parameters have been standardized assuming the space group of lowest symmetry in agreement with the Pearson symbol, e.g. $P222$ for oP^* or o^{**} .

Standardized data are described with respect to the standard settings described in *International Tables for Crystallography* Volume A, with the following additional restraints: inversion centre at the origin, unique b axis and 'best' cell for monoclinic structures (Parthé & Gelato, 1985), triple-hexagonal cell for rhombohedral structures or Niggli reduced cell for triclinic structures. As a consequence, they can easily be incorporated into any program handling crystallographic data. The systematic standardization of the crystallographic data also greatly simplifies the classification of crystal structures into different prototypes.

A conversion tool to standardize cell parameters and/or compute the Niggli reduced cell is included in the software of Pearson's Crystal Data.

3.7.5.4. Consequent prototype assignment

The prototype is a well known concept in inorganic chemistry, where a large number of compounds often crystallize with very similar atom arrangements. The compilation *Strukturbericht* started to catalogue crystal structures into types named by codes such as A1, B1 or A15. These notations are still in use; however, today prototypes are generally referred to by the name of the compound for which this particular kind of atom arrangement was first identified, i.e. Cu, NaCl and Cr_3Si for the types enumerated above. Pearson's Crystal Data uses a longer notation which also includes the Pearson symbol and the space-group number: Cu, $cF4,225$, NaCl, $cF8,225$ and $\text{Cr}_3\text{Si},cP8,223$. In a few cases several prototypes correspond to the same code, for example several polytypes of CdI_2 have the same notation. A similar situation occurs for the wrong and the correct structure proposals for FeB, which have the same Pearson code and space group. In these cases a letter is added after the type-defining compound, for example the correct FeB type will be referred to as FeB-b, $oP8,62$.

Each prototype is defined on a particular PCD database entry. In principle, this data set represents a recent refinement of the structure of the type-defining compound, but no effort has been made to find or use the most recent determination.

All of the data sets with published coordinates in Pearson's Crystal Data have been classified into prototypes following the criteria defined in *TYPIX* (Parthé *et al.*, 1993, 1994). According to this definition, isotypic compounds must crystallize in the same space group and have similar cell-parameter ratios; the atoms should occupy the same Wyckoff positions in the standardized description and have similar positional coordinates. If all of these criteria are fulfilled, the atomic environments should be similar. Note that H^+ (protonic hydrogen) is ignored in the assignment of the prototype as well as in the Wyckoff sequence, Pearson symbol/code and atomic environments. Isopointal substitution variants are usually distinguished; however, no distinction is made between structures with fully and partly occupied atom sites. At present, 29 470 prototypes are represented.

When possible, a prototype has also been assigned to data sets without published atom coordinates. The prototype is often stated in the publication; in other cases the editors have assigned it. The editor will have added the exact space-group setting to which the cell parameters refer when this was not published. It is important to note that a prototype has been assigned at two different levels. The first is intimately related to the published

data (entry level), whereas the second is assigned at the phase level and may, in some cases, be inconsistent with the crystallographic data listed below.

For partly investigated structures, the available structural information is given using a similar way, for example the complete Pearson symbol may be replaced by t^{**} (tetragonal) or cI^* (cubic body-centred) and the place of the type-defining compound is occupied by an asterisk.

3.7.5.5. Assigned atom coordinates

In order to give an approximate idea of the actual structure, a complete set of positional coordinates and site occupancies is proposed for data sets where a prototype could be assigned but the atom coordinates were not determined. The coordinates of the type-defining entry are proposed as a first approximation. The atom distribution is inserted by an *ESDD* module that compares the chemical formula of the type-defining entry with the chemical formula of the isotypic compound where the chemical elements have been reordered by the editor so that the first element is expected to occupy the same atom sites as the first element in the type-defining formula, and so on. Depending on the character of the prototype, substitutions and/or vacancies are either distributed over all atom sites occupied by the corresponding element or are expected to occur selectively on particular atom sites.

For this category of database entries, structure drawings, diffraction patterns and interatomic distances have also been computed. The structural portion of the database is thus more extensive than the primary literature.

3.7.5.6. External links

When relevant, the database entries contain links to external data sources, including ASM International Alloys Phase Diagrams Centre Online, SpringerMaterials (The Landolt-Börnstein Database incorporating Inorganic Solid Phases PAULING FILE Multinaries Edition – 2010 in SpringerMaterials) and the original publication (through <https://www.crossref.org/>). A (static) reference to the Powder Diffraction File entry number is provided for database entries that are included in the PDF4+ product.

3.7.5.7. Retrievable database fields

In addition to bibliographic (e.g. a particular institute) and chemical (e.g. sulfates) searches, many characteristics of the experiment and data processing (e.g. single crystal, neutron diffraction, range of temperature or reliability factors) or additional studies (e.g. pressure-dependence studies, magnetic structure) can be used as search criteria. Published crystal data, standardized crystal data and the Niggli reduced cells can be searched, as well as crystallographic classifications such as crystal class, Pearson symbol, Pearson code, Wyckoff sequence, structure prototype or structure class. Such searches can be very valuable in identifying a structural model for a new composition and saving the work of an *ab initio* structure determination.

The Quick Search pane includes commonly used searches on chemical elements (including cations in a particular oxidation state for oxides and halides), the number of elements and functional groups. The chemical selection (and/or/not) can be combined with selection on structure prototypes, space-group numbers and symbols or the crystal system. Retrieval on cell parameters (with ranges) and bibliographic information is also