

## 3. CIF DATA DEFINITION AND CLASSIFICATION

Table 3.3.3.1. Category groups defined in the powder CIF dictionary

The groups are listed in the order in which they are described in this chapter.

Section	Category group	Subject covered
<i>(a) Experimental measurements</i>		
3.3.4.1	PD_CHAR	Characterization of a sample
3.3.4.1	PD_PREP	Preparation of a sample
3.3.4.2	PD_SPEC	Specimen used in an experiment
3.3.4.3	PD_CALIB	Calibration information
3.3.4.3	PD_INSTR	The experimental instrument
3.3.4.4	PD_MEAS	Raw measurements and instrumental settings
<i>(b) Analysis</i>		
3.3.5.1	PD_PROC	Processed settings
3.3.5.2	PD_CALC	Simulated settings
3.3.5.3	PD_PEAK	Diffraction peak table
3.3.5.4	REFLN	Reflection assignments and intensities
<i>(c) Atomicity, chemistry and structure</i>		
3.3.6.1	PD_PHASE	Phases present
<i>(d) File metadata</i>		
3.3.7.1	PD_BLOCK	Relationships between data blocks
3.3.4.3, 3.3.4.4, 3.3.5.1, 3.3.5.2	PD_DATA	Measured and simulated intensities

separate loops, for example, if the increment between data points differs.

This need to contain diverse items in a common ‘looped’ list has led the pdCIF dictionary to use category names in a different way from the other CIF dictionaries, in which CIF data items are usually named according to their category. In the pdCIF dictionary, data items that might appear in the loop for diffraction intensities are assigned to the category PD\_DATA. Only one data item is named using this category as prefix, `_pd_data_point_id`. Another departure from the convention used in other dictionaries is that several `_pd_refl*_` data names are assigned to the category REFLN so that these items may be included in a loop with `_refln_*` items defined in the core CIF dictionary.

Table 3.3.3.1 summarizes the category groups in the pdCIF dictionary; the individual categories are listed alphabetically in Appendix 3.3.1. The appendix also lists for each category the section of this chapter in which the category is described.

The order in which the categories are discussed follows the scheme of Table 3.1.10.1, so that the contents of the dictionary are summarized under the headings *Experimental measurements* (Section 3.3.4), *Analysis* (Section 3.3.5), *Atomicity, chemistry and structure* (Section 3.3.6) and *File metadata* (Section 3.3.7). The pdCIF dictionary does not contribute any new data items relevant to publication beyond those already in the core CIF dictionary.

The data items in each category are listed below. Category keys, if specified, are listed first and are marked by a bullet (●); the remaining data items in each category are listed alphabetically. Note that the category PD\_DATA is discussed in several different sections.

### 3.3.4. Experimental measurements

The categories in the powder CIF dictionary relating to the crystallographic experiment are as follows:

#### *Characterization and preparation of the sample* (§3.3.4.1)

PD\_CHAR group  
 PD\_CHAR  
 PD\_PREP group  
 PD\_PREP

#### *Description of the specimen* (§3.3.4.2)

PD\_SPEC group  
 PD\_SPEC

#### *Instrument calibration and design* (§3.3.4.3)

PD\_CALIB group  
 PD\_CALIB  
 PD\_CALIBRATION  
 PD\_INSTR group  
 PD\_INSTR

#### *Observations and measurement conditions* (§3.3.4.4)

PD\_DATA group  
 PD\_DATA (items beginning with `_pd_meas_*`)  
 PD\_MEAS group  
 PD\_MEAS\_INFO  
 PD\_MEAS\_METHOD

The pdCIF dictionary differentiates between the terms *sample* and *specimen*. The terms are often treated as interchangeable, but they have quite distinct meanings. The term *sample* refers to a batch of material, while the term *specimen* refers to the particular portion of the sample that was used for a measurement. In some cases, the specimen is modified before it is used for data collection. For example, it may be mixed with an internal standard, dried, hydrated or pressed into a pellet.

#### 3.3.4.1. Characterization and preparation of the sample

The data items in these categories are as follows:

- (a) PD\_CHAR  
 ● `_pd_char_atten_coef_mu_calc`  
 ● `_pd_char_atten_coef_mu_obs`  
 ● `_pd_char_colour`  
 ● `_pd_char_particle_morphology`  
 ● `_pd_char_special_details`
- (b) PD\_PREP  
 ● `_pd_prep_conditions`  
 ● `_pd_prep_cool_rate`  
 ● `_pd_prep_pressure`  
 ● `_pd_prep_temperature`

The PD\_CHAR data items describe information known about the sample from observation and chemical analysis. For example, a description of the sample morphology can be specified using `_pd_char_particle_morphology`. Note that there are data items in the core dictionary that are appropriate for use with powder diffraction. For example, `_atom_type_analytical_mass_%` can be used for chemical analysis results and `_chemical_melting_point` for the melting point. Several similar data items occur in the pdCIF and core dictionaries. `_expt1_crystal_colour` and `_pd_char_colour` both describe the sample colour, but `_pd_char_colour` is more systematic. Also, `_pd_char_atten_coef_mu_calc` and `_expt1_absorpt_coefficient_mu` describe similar properties, but `_pd_char_atten_coef_mu_calc` is adjusted for the sample packing fraction, so it can be compared with the experimental value, `_pd_char_atten_coef_mu_obs`, when a direct measurement is made.

The PD\_PREP data items describe how the sample was collected or prepared. For example, `_pd_prep_pressure` and `_pd_prep_temperature` describe the pressure and temperature used to prepare the sample. Note that these will probably differ from the pressure and temperature conditions at which diffraction measurements are made. Measurement conditions are recorded in `_diffrn_ambient_pressure` and `_diffrn_ambient_temperature`.

## 3.3.4.2. Description of the specimen used in the experiment

The data items in this category are as follows:

```
PD_SPEC
  _pd_spec_description
  _pd_spec_mount_mode
  _pd_spec_mounting
  _pd_spec_orientation
  _pd_spec_preparation
  _pd_spec_shape
  _pd_spec_size_axial
  _pd_spec_size_equat
  _pd_spec_size_thick
  _pd_spec_special_details
```

The PD\_SPEC data items describe the specimen used to measure the diffraction data. The data item `_pd_spec_preparation` describes how the specimen that was used to measure the diffraction data was treated, not how the sample was prepared (PD\_PREP) or characterized (PD\_CHAR).

The PD\_SPEC data items are also used to describe how the specimen was mounted for the diffraction experiment. For example, `_pd_spec_mount_mode` and `_pd_spec_orientation` describe the measurement geometry, while `_pd_spec_shape` and `_pd_spec_size_*` describe the specimen shape and size.

## 3.3.4.3. Instrument calibration and description

The data items in these categories are as follows:

## (a) PD\_CALIB

```
• _pd_calib_detector_id
  _pd_calib_2theta_offset
  _pd_calib_2theta_off_point
  _pd_calib_2theta_off_min
  _pd_calib_2theta_off_max
  _pd_calib_detector_response
  _pd_calib_std_external_block_id
  _pd_calib_std_external_name
  _pd_calib_std_internal_mass_%
  _pd_calib_std_internal_name
```

## (b) PD\_CALIBRATION

```
_pd_calibration_conversion_eqn
_pd_calibration_special_details
```

## (c) PD\_INSTR

```
_pd_instr_2theta_monochr_pre
_pd_instr_2theta_monochr_post
_pd_instr_beam_size_ax
_pd_instr_beam_size_eq
_pd_instr_cons_illum_flag
_pd_instr_cons_illum_len
_pd_instr_dist_src/mono
_pd_instr_dist_mono/spec
_pd_instr_dist_src/spec
_pd_instr_dist_spec/anal
_pd_instr_dist_anal/detc
_pd_instr_dist_spec/detc
_pd_instr_divg_ax_src/mono
_pd_instr_divg_ax_mono/spec
_pd_instr_divg_ax_src/spec
_pd_instr_divg_ax_spec/anal
_pd_instr_divg_ax_anal/detc
_pd_instr_divg_ax_spec/detc
_pd_instr_divg_eq_src/mono
_pd_instr_divg_eq_mono/spec
_pd_instr_divg_eq_src/spec
_pd_instr_divg_eq_spec/anal
_pd_instr_divg_eq_anal/detc
_pd_instr_divg_eq_spec/detc
_pd_instr_geometry
_pd_instr_location
_pd_instr_monochr_pre_spec
_pd_instr_monochr_post_spec
_pd_instr_slit_ax_src/mono
_pd_instr_slit_ax_mono/spec
_pd_instr_slit_ax_src/spec
_pd_instr_slit_ax_spec/anal
_pd_instr_slit_ax_anal/detc
```

```
_pd_instr_slit_ax_spec/detc
_pd_instr_slit_eq_src/mono
_pd_instr_slit_eq_mono/spec
_pd_instr_slit_eq_src/spec
_pd_instr_slit_eq_spec/anal
_pd_instr_slit_eq_anal/detc
_pd_instr_slit_eq_spec/detc
_pd_instr_soller_ax_src/mono
_pd_instr_soller_ax_mono/spec
_pd_instr_soller_ax_src/spec
_pd_instr_soller_ax_spec/anal
_pd_instr_soller_ax_anal/detc
_pd_instr_soller_ax_spec/detc
_pd_instr_soller_eq_src/mono
_pd_instr_soller_eq_mono/spec
_pd_instr_soller_eq_src/spec
_pd_instr_soller_eq_spec/anal
_pd_instr_soller_eq_anal/detc
_pd_instr_soller_eq_spec/detc
_pd_instr_source_size_ax
_pd_instr_source_size_eq
_pd_instr_special_details
```

## (d) Part of PD\_DATA

```
_pd_instr_var_illum_len
```

The bullet (•) indicates a category key.

Calibration information can be placed in the PD\_CALIB and PD\_CALIBRATION categories. The `_pd_calibration_*` data items are descriptive and will not appear in a loop. The `_pd_calib_*` items may be looped to describe multiple detectors. Correction values for  $2\theta$  can be given using the `_pd_calib_2theta_offset` and `_pd_calib_2theta_off_*` data items. A calibration equation can be given using `_pd_calibration_conversion_eqn`. When multiple detectors are used, `_pd_calib_detector_response` is used to indicate the relative performance of each detector. The detector deadtime is specified using the core data item `_diffrn_detector_dtime` (which cannot be looped by detector).

If an internal standard is added to the sample for calibration, this information is specified using `_pd_calib_std_internal_name` to specify the material added and `_pd_calib_std_internal_mass_%` to specify the amount.

When a set of calibration intensities is measured using an external standard, it is possible to include the measurements and the derived results in a separate CIF block. A data block would then use `_pd_calib_std_external_block_id` to link to the block containing the calibration information. See Section 3.3.7 for a discussion of block pointers and block IDs. Note that the use of a unique name for the block ID allows the calibration information to be stored in a separate file, so that the calibration CIF need not be repeated in every CIF that references it.

The PD\_INSTR section of the pdCIF dictionary contains terms that describe the instrument used. For example, the instrument or laboratory location is given using `_pd_instr_location`. The instrument type can be indicated using `_pd_instr_geometry`. The instrument geometry can be described in much greater detail using several data items. The geometry is described in terms of four regions of the experiment: radiation source to monochromator (`src/mono`); monochromator to specimen (`mono/spec`); specimen to analyser (`spec/anal`); and analyser to detector (`anal/detc`). If no monochromator is present, the first two regions are combined into radiation source to specimen (`src/spec`). If no analyser is present, the last two regions are combined into specimen to detector (`spec/detc`). Thus two, three or four sets of values describe the dimensions of the instrument and the collimation. For example, `_pd_instr_dist_src/mono` would be used to specify the distance between the radiation source and the monochromator. Alternatively, `_pd_instr_dist_src/spec` would be used to specify the distance between the

### 3. CIF DATA DEFINITION AND CLASSIFICATION

radiation source and the specimen if no monochromator was present.

Two methods may be used to describe the slits limiting the divergence in the equatorial plane. The angular divergence allowed by the slits in degrees can be specified using the `_pd_instr_divg_eq_*` data items. Alternatively, the dimensions of the slits in the equatorial direction in millimetres may be specified using `_pd_instr_slit_eq_*`. The dimensions of these slits in the axial direction, *i.e.* the direction perpendicular to the equatorial plane and containing the incident or diffracted beam as appropriate, are specified in millimetres using `_pd_instr_slit_ax_*`. The axial slit lengths, along with the `_pd_instr_dist_*` distances, are useful for estimating the low-angle peak asymmetry (Finger *et al.*, 1994). Note that angular divergence in the axial plane is not a well defined concept for line-focus instruments, but can be specified, where appropriate, using `_pd_instr_divg_ax_*`.

The axial and equatorial directions are shown schematically in Fig. 3.3.4.1. The equatorial plane contains the equatorial direction vectors, as well as the incident beam, the diffracted beam and the scattering vector. The axial plane is perpendicular to the equatorial plane and contains the sample centre, which is the point where the incident and diffracted beams meet. For area-detection instruments, the designations of axial and equatorial directions may be arbitrary.

Soller collimators are described using `_pd_instr_soller_eq_*` data items rather than `_pd_instr_divg_eq_*` data items. It is common practice to specify the Soller collimation in arc-minutes (*e.g.* 30'). However, pdCIF defines these items to have units of degrees, so 30' would be recorded in the CIF as 0.5. It is not usual to limit the axial divergence, except to reduce low-angle asymmetry, but if this is done, the `_pd_instr_soller_ax_*` data items can be used to define this.

For constant-wavelength instruments, it is common to have a monochromator or filter either before the sample, or after the sample (an analyser), or sometimes both. This is described using `_pd_instr_monochr_pre_spec` and `_pd_instr_monochr_post_spec`. It is rare, but possible, to have both a filter and a monochromator in the same location. The BT-1 neutron powder diffractometer at NIST uses both a Cu(311) monochromator and a graphite filter to attenuate the  $\lambda/2$  component. In this case, the two elements would be placed in a loop:

```
loop_ _pd_instr_monochr_pre_spec
  '5 seg. vert. focusing Cu(311) monochromator'
  '3 cm graphite filter'
```

Note that the monochromator and analyser takeoff angles are given using `_pd_instr_2theta_monochr_pre` and `_pd_instr_2theta_monochr_post`. It is useful to record these values for X-ray studies, as they are needed for proper polarization corrections.

In a conventional Bragg–Brentano diffractometer, the divergence slits limit the illumination area at the sample. However, since the  $\varphi$  axis (the sample  $\theta$  axis) is usually set to bisect the  $2\theta$  angle of the detector, the actual length of the area of the sample that is illuminated changes with  $2\theta$ . One should choose divergence slits so that the beam does not illuminate areas outside the sample at the lowest diffraction angle used. An alternative method for data collection is to have a divergence slit that opens as  $2\theta$  increases, so that a constant area of the sample is illuminated. This is known as a  $\theta$ -compensating slit. Using a  $\theta$ -compensating slit provides a better signal-to-noise ratio at larger  $2\theta$  values, but means that the diffraction intensities have to be normalized to compensate for the change in illumination. It also introduces greater optical

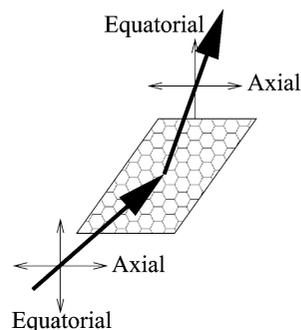


Fig. 3.3.4.1. The axial and equatorial directions in a powder-diffraction experiment.

aberrations with  $2\theta$ ; a flat plate becomes an increasingly worse approximation to the curved sample geometry in the true Bragg–Brentano geometry.

The use of a variable divergence slit can be recorded in the form:

```
_pd_instr_cons_illum_flag    yes
_pd_instr_cons_illum_len    25.4
```

Note that if `_pd_instr_cons_illum_flag` is not specified, the value is assumed to be `no`, indicating that a fixed-width divergence slit has been used.

The beam size can be specified in two different ways: as the size at the source, using `_pd_instr_source_size_ax` and `_pd_instr_source_size_eq`, or as the size at the sample position, using `_pd_instr_beam_size_ax` and `_pd_instr_beam_size_eq`. Note that the size of the beam at the sample differs from the illumination length described above except when the sample is perpendicular to the beam. When a variable-divergence slit is in use, the beam size at the sample changes with  $2\theta$ , so if this size is known directly, the `_pd_instr_beam_size_eq` data item can be included in the loop containing the diffraction intensities. Similarly, in a constant-divergence instrument, where the illumination length changes with  $2\theta$ , the illumination length can be specified in the loop using `_pd_instr_var_illum_len`.

There are also several data items in the core CIF dictionary that should be present in the description of the instrument in a pdCIF. Use `_diffrn_radiation_probe` and `_diffrn_radiation_type` to specify the type of radiation used and `_diffrn_detector_type` to specify the detection type.

#### 3.3.4.4. Observations and measurement conditions

The data items in these categories are as follows:

- (a) Part of PD\_DATA
- ```
_pd_data_point_id
_pd_meas_2theta_fixed
_pd_meas_2theta_range_min
_pd_meas_2theta_range_max
_pd_meas_2theta_range_inc
_pd_meas_2theta_scan
_pd_meas_angle_2theta
_pd_meas_angle_chi
_pd_meas_angle_omega
_pd_meas_angle_phi
_pd_meas_counts_total
_pd_meas_counts_background
_pd_meas_counts_container
_pd_meas_counts_monitor
_pd_meas_datetime_initiated
_pd_meas_detector_id
  → _pd_calib_detector_id
_pd_meas_point_id
_pd_meas_position
_pd_meas_rocking_angle
_pd_meas_step_count_time
_pd_meas_time_of_flight
```

### 3.3. CLASSIFICATION AND USE OF POWDER DIFFRACTION DATA

(b) PD\_MEAS\_INFO  
\_pd\_meas\_info\_author\_address  
\_pd\_meas\_info\_author\_email  
\_pd\_meas\_info\_author\_fax  
\_pd\_meas\_info\_author\_name  
\_pd\_meas\_info\_author\_phone  
\_pd\_meas\_intensity\_total  
\_pd\_meas\_intensity\_background  
\_pd\_meas\_intensity\_container  
\_pd\_meas\_intensity\_monitor

(c) PD\_MEAS\_METHOD  
\_pd\_meas\_number\_of\_points  
\_pd\_meas\_rocking\_axis  
\_pd\_meas\_scan\_method  
\_pd\_meas\_special\_details  
\_pd\_meas\_units\_of\_intensity

The arrow (→) is a reference to a parent data item.

The item `_pd_data_point_id` identifies each entry in the list of measured, processed or simulated intensities. It is the only item in the PD\_DATA category that actually begins with the string `_pd_data_`, and is included here for convenience. If the list of intensities is split across several distinct loops, the role of this identifier may be adopted by other identifiers, such as `_pd_meas_point_id` in an isolated list of measured intensities.

The `_pd_meas_*` data items contain unprocessed measurements and documentation on the instrumental settings used for the measurements. Note that the choice of the data items used to represent this information is determined by the type of diffraction instrument, as well as how the measurement was conducted. This will be discussed further in Section 3.3.8. However, some `_pd_meas_*` data items are independent of the instrument type. For example, the use of `_pd_meas_datetime_initiated` is good practice, as is use of the `_pd_meas_info_author_*` data items. It is probably good practice to record the number of data points in `_pd_meas_number_of_points` for the benefit of people who might read the CIF, but there is no requirement that this item be present. This means that software should determine the number of data points directly when reading the CIF, rather than relying on the presence of a value for `_pd_meas_number_of_points`.

#### 3.3.5. Analysis

The categories relating to the information derived from the measurements are as follows:

*Processed intensities, positions and data processing* (§3.3.5.1)

PD\_DATA group  
PD\_DATA (items beginning with `_pd_proc_*`)

PD\_PROC group  
PD\_PROC\_INFO  
PD\_PROC\_LS

*Simulated intensities and their positions* (§3.3.5.2)

PD\_DATA group  
PD\_DATA (items beginning with `_pd_calc_*`)

PD\_CALC group  
PD\_CALC

*Diffraction peak table* (§3.3.5.3)

PD\_PEAK group  
PD\_PEAK  
PD\_PEAK\_METHOD

*Reflection assignments and intensities* (§3.3.5.4)

REFLN

In Rietveld and other studies, processed or simulated intensities are presented alongside measured values. This leads to the presence of both derived and measured values in the same category (PD\_DATA). However, the purposes of the data items that

refer to processed and simulated data points are made clear by the way they are named. Overall descriptions of processed and simulated intensity data are covered by the categories PD\_PROC\_INFO, PD\_PROC\_LS and PD\_CALC. The two categories PD\_PEAK and PD\_PEAK\_METHOD are used to describe lists of peak positions, which would typically be used to search and match powder profiles. Some additional data items relevant to the table of Bragg reflections are defined as additions to the existing REFLN category in the core CIF dictionary.

#### 3.3.5.1. Processed intensities, their positions and processing information

The data items in these categories are as follows:

(a) Part of PD\_DATA  
\_pd\_proc\_2theta\_corrected  
\_pd\_proc\_2theta\_range\_min  
\_pd\_proc\_2theta\_range\_max  
\_pd\_proc\_2theta\_range\_inc  
\_pd\_proc\_d\_spacing  
\_pd\_proc\_energy\_incident  
\_pd\_proc\_energy\_detection  
\_pd\_proc\_intensity\_net  
\_pd\_proc\_intensity\_total  
\_pd\_proc\_intensity\_bkg\_calc  
\_pd\_proc\_intensity\_bkg\_fix  
\_pd\_proc\_intensity\_incident  
\_pd\_proc\_intensity\_norm  
\_pd\_proc\_point\_id  
\_pd\_proc\_recip\_len\_Q  
\_pd\_proc\_wavelength

(b) PD\_PROC\_INFO  
\_pd\_proc\_info\_author\_address  
\_pd\_proc\_info\_author\_email  
\_pd\_proc\_info\_author\_fax  
\_pd\_proc\_info\_author\_name  
\_pd\_proc\_info\_author\_phone  
\_pd\_proc\_info\_data\_reduction  
\_pd\_proc\_info\_datetime  
\_pd\_proc\_info\_excluded\_regions  
\_pd\_proc\_info\_special\_details

(c) PD\_PROC\_LS  
\_pd\_proc\_ls\_background\_function  
\_pd\_proc\_ls\_peak\_cutoff  
\_pd\_proc\_ls\_pref\_orient\_corr  
\_pd\_proc\_ls\_prof\_R\_factor  
\_pd\_proc\_ls\_prof\_wR\_factor  
\_pd\_proc\_ls\_prof\_wR\_expected  
\_pd\_proc\_ls\_profile\_function  
\_pd\_proc\_ls\_special\_details  
\_pd\_proc\_ls\_weight

The pdCIF dictionary distinguishes between values that are measured directly and values that are derived from these observations. For example, in a constant-wavelength instrument, diffraction intensities are recorded as a function of  $2\theta$ . One may derive  $d$ -space values from the  $2\theta$  values using the value of the wavelength and corrections for the  $2\theta$  zero-point error and the sample displacement. One may also derive a new set of data points from the observations, for example by summing adjacent data points when the increment between the data points is much smaller than is warranted by the peak widths. For peak searching and other non-quantitative purposes, the diffraction intensities may be smoothed or otherwise modified. Note that the unprocessed measurement values are retained using the data items `_pd_meas_*`. Since the original measurements are still available, modifications like these do not result in the loss of the original data. In fact, by placing processed values in multiple blocks, a single CIF may contain measurements that have been processed in more than one way.

It is good practice to use the `_pd_proc_info_author_*` and `_pd_proc_info_datetime` data items. It is also a good idea to