

## 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

### 3. CIF DATA DEFINITION AND CLASSIFICATION

describe how the measurements were processed using `_pd_proc_info_data_reduction`.

The `_pd_proc_*` data items in this list may be used to calibrate the  $2\theta$  or energy values of the data. These are defined in the items `_pd_proc_2theta_corrected`, `_pd_proc_2theta_range_*`, `_pd_proc_d_spacing`, `_pd_proc_energy_*`, `_pd_proc_recip_len_Q` and `_pd_proc_wavelength`.

When corrections, scaling or other processing, such as averaging or smoothing, are applied to the intensities, the results are stored using the `_pd_proc_intensity_*` data items. Note that if the number of data points does not change, it might be most convenient to include the processed intensities in the same loop as the observed values. This is not always possible, so these items can be placed in a separate loop if there is no longer a one-to-one correspondence between the  $2\theta$  or energy positions for the `_pd_proc_intensity_*` values and the `_pd_meas_counts_*` or `_pd_meas_intensity_*` values.

For energy-dispersive measurements, the incident spectrum must be determined for normalization. This can be recorded using `_pd_proc_intensity_incident`. For other types of normalization, `_pd_proc_intensity_norm` should be used.

For full-pattern fitting, there is a series of `_pd_proc_ls_*` data items for recording settings and results. For example, agreement factors can be recorded using the `_pd_proc_ls_prof_*_factor` data items. Some data items may be included in the loop(s) containing the measured or the processed data: `_pd_proc_ls_weight` specifies the weight assigned to each point and `_pd_proc_intensity_bkg_calc` specifies the fitted background. Note that background values are usually generated by extrapolation from fixed values set during the refinement or are determined from a function that is fitted to the observations, and occasionally both are used together. The function that has been fitted can be described using `_pd_proc_ls_background_function`, while fixed points are listed using `_pd_proc_intensity_bkg_fix`. If sections of the pattern are not fitted, this is indicated using `_pd_proc_info_excluded_regions`.

#### 3.3.5.2. Simulated data

The data items in these categories are as follows:

(a) Part of PD\_DATA  
`_pd_calc_intensity_net`  
`_pd_calc_intensity_total`  
`_pd_calc_point_id`

(b) PD\_CALC  
`_pd_calc_method`

It is common to calculate powder-diffraction intensities from a crystallographic model. This is necessary for Rietveld refinements, where the model is fitted to the experimentally observed intensities. It is also used to simulate the diffraction pattern of a material for which the structure is known, perhaps for comparison with a measured diffraction pattern.

A crystallographic model can be described in CIF using data items from the core CIF dictionary, as described in Chapter 3.2. To record the results of the simulation, the data items `_pd_calc_intensity_net` or `_pd_calc_intensity_total` are used. The difference between these two data items depends on the treatment of background. If the pattern is simulated with a fitted background added to it, `_pd_calc_intensity_total` is used; otherwise `_pd_calc_intensity_net` is used. The values will typically be placed in a loop with the processed (`_pd_proc_*`) data items or the observed (`_pd_meas_*`) data items. If neither observed

nor processed data are present (e.g. for a simulation), or if, for some reason, the simulation has been performed with a different  $2\theta$  range or step size, the appropriate `_pd_proc_*` data items are used to define the  $2\theta$  values *etc.* used for the simulation.

#### 3.3.5.3. Diffraction peak table

The data items in these categories are as follows:

(a) PD\_PEAK  
• `_pd_peak_id`  
`_pd_peak_2theta_centroid`  
`_pd_peak_2theta_maximum`  
`_pd_peak_d_spacing`  
`_pd_peak_intensity`  
`_pd_peak_pk_height`  
`_pd_peak_wavelength_id`  
→ `_diffrn_radiation_wavelength_id`  
`_pd_peak_width_2theta`  
`_pd_peak_width_d_spacing`

(b) PD\_PEAK\_METHOD  
`_pd_peak_special_details`

The bullet (•) indicates a category key. The arrow (→) is a reference to a parent data item. Items in italics are defined in the core CIF dictionary.

When diffraction intensities are first measured, particularly when attempting to identify unknown phases in a material, the first step in the analysis is often to compile a list of peak positions. These peak positions are commonly used to search the Powder Diffraction File, which contains lists of peak heights and positions for approximately 100 000 materials (International Centre for Diffraction Data, 2004).

Information on diffraction peaks is recorded in the PD\_PEAK section of the pdCIF. Peak positions are recorded using `_pd_peak_2theta_maximum` or `_pd_peak_2theta_centroid`, for positions determined from the intensity maxima or from the peak centroids, respectively. It is also possible to record peak positions using `_pd_peak_d_spacing`. Peak intensities are recorded using `_pd_peak_intensity` and `_pd_peak_pk_height`, for the integrated peak area or the intensity value at the peak maximum, respectively. Peak widths are recorded using `_pd_peak_width_2theta` and `_pd_peak_width_d_spacing`.

A separate loop is used to list reflections, as will be discussed in Section 3.3.5.4. To link reflections to peaks (one peak may consist of many reflections), each peak is assigned a unique code using `_pd_peak_id`, which is then referenced in the reflection table using `_pd_refl_peak_id`.

When intensities are measured using radiation with more than one wavelength, for example when both  $\text{Cu } K\alpha_1$  and  $K\alpha_2$  radiation are used or when a monochromator passes both  $\lambda$  and  $\lambda/2$  radiation, peaks may be assigned a wavelength symbol using `_pd_peak_wavelength_id`, where the wavelength symbol is defined in a separate `_diffrn_radiation_wavelength_id` loop. However, for many experiments, the assignment of wavelengths to peaks will be impractical owing to reflection overlap. It is usually better practice to specify wavelength labels in the reflection table using `_pd_refl_wavelength_id`.

#### 3.3.5.4. Reflection assignments and intensities

In addition to the REFLN data items defined in the core CIF dictionary, the following items are defined:

REFLN  
`_pd_refl_peak_id`  
→ `_pd_peak_id`  
`_pd_refl_phase_id`  
→ `_pd_phase_id`