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 ($\rightarrow$) *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 2θ. One may derive d-space values from the 2θ values using the value of the wavelength and corrections for the 2θ 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

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 2θ 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 2θ 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 2θ range or step size, the appropriate `_pd_proc_*` data items are used to define the 2θ 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_refln_peak_id`.

When intensities are measured using radiation with more than one wavelength, for example when both Cu $K\alpha_1$ and $K\alpha_2$ radiation are used or when a monochromator passes both λ 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_refln_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_refln_peak_id
        → _pd_peak_id
   _pd_refln_phase_id
        → _pd_phase_id
```

† `_pd_refln_wavelength_id` (`_refln_wavelength_id`)
        → `_diffrn_radiation_wavelength_id`

The arrow ($\rightarrow$) is a reference to a parent data item. The dagger (†) indicates a deprecated item, which should not be used in the creation of new CIFs. Items in italics are defined in the core CIF dictionary.

In a single-crystal experiment, a reflection table contains the initial experimental observations for structural analysis. In contrast, the reflection table for a powder-diffraction experiment is a derived result that depends on the model used to apportion intensity between overlapping reflections. Another difference is that in a single-crystal experiment, the reflection list will refer to only one phase (one hopes), while it is common to have reflections from more than one phase in a powder-diffraction reflection list.

A list of reflections in a powder-diffraction pattern is commonly generated by Rietveld analysis, where Hugo Rietveld's algorithm (Rietveld, 1967, 1969) is used to estimate the intensity of each reflection. Alternatively, when the structure of one or more phases is not known, it is possible to use full-pattern intensity-extraction methods such as the algorithms developed by Pawley (1981) or Le Bail *et al.* (1988). In fact, intensity information obtained by full-pattern intensity extraction is often used for *ab initio* structure determination.

Most of the information in the reflection table will be defined using data items from the core CIF dictionary (see Section 3.2.2.2 and Chapter 4.1). For example, `_refln_index_h`, `_refln_index_k` and `_refln_index_l` will be used for the indices. The structure factors and reflection intensities are specified using `_refln_intensity_calc`, `_refln_intensity_meas`, `_refln_F_squared_calc` and `_refln_F_squared_meas`; reflection positions are defined using `_refln_d_spacing`. To link a reflection with a powder-diffraction peak, the pdCIF data item `_pd_refln_peak_id` is used. The value for `_pd_refln_peak_id` serves as a pointer to an entry in the peak table which has been labelled, using the data name `_pd_peak_id`, with the same symbol. Likewise, to link a reflection to a phase, the pdCIF data item `_pd_refln_phase_id` points to a phase defined using `_pd_phase_id` in the phase table. Since a single reflection may be observed with more than one wavelength, for example, with $\lambda/2$ or $K\alpha_2$ wavelengths, the pdCIF dictionary defines a wavelength link, `_pd_refln_wavelength_id`, that defines a wavelength label. However, since version 2.1, the core CIF dictionary defines `_refln_wavelength_id` and this should be used in preference to `_pd_refln_wavelength_id`. The data items `_refln_wavelength_id` and `_pd_refln_wavelength_id` both point to a wavelength label defined using `_diffrn_radiation_wavelength_id`.

The International Centre for Diffraction Data abstracts peak positions and heights for inclusion in the Powder Diffraction File. This information would be found in the `_pd_peak` section of a pdCIF. However, in many studies, particularly in Rietveld refinements, peak tables are never generated. In principle, it should be possible to calculate peak positions and peak heights (or better still, peak areas) from the information in a reflection table. An algorithm for this would be very useful.

3.3.6. Atomicity, chemistry and structure

The structural model of a compound determined by powder-diffraction methods can be described by the data items in the core CIF dictionary. However, for a powder-diffraction study of a mixture of phases, the PD_PHASE category is used to list the phases present. This is the only category in the pdCIF dictionary that extends the description of the structural model beyond that covered by items in the core CIF dictionary.

3.3.6.1. Table of phases

The data items in this category are as follows:

PD_PHASE
● `_pd_phase_id`
  `_pd_phase_block_id`
  `_pd_phase_mass_%`
  `_pd_phase_name`

The bullet (●) indicates a category key.

When a sample contains more than one phase, the PD_PHASE data items are used to create a table describing the phases present. For example, the name and abundance of each phase can be specified using `_pd_phase_name` and `_pd_phase_mass_%`, respectively.

Two types of pointers can also be defined:

(i) Since the crystallographic description of each phase must be incorporated in a separate data block, `_pd_phase_block_id` contains the unique block ID (see Section 3.3.7) pointing to the block containing the data for the phase.

(ii) An arbitrary label is assigned to every phase using `_pd_phase_id` so that reflections can be assigned to a phase using `_pd_refln_phase_id`. This is discussed further in Section 3.3.5.4.

3.3.7. File metadata

The many data items in the core dictionary that describe file auditing and history cover most of the metadata requirements of a pdCIF, but two new data items in the pdCIF category PD_BLOCK are introduced to provide a specific mechanism for identifying and relating individual data blocks.

Data items in this category are as follows:

PD_BLOCK
  `_pd_block_diffractogram_id`
  `_pd_block_id`

The data item `_pd_block_id` is used to define a unique name for each data block. This name is used so that one data block may reference another data block. Since CIF blocks may be separated into different files, or many CIFs from different sources may be grouped into a single file, the block ID provides a robust mechanism for maintaining references between blocks, independent of how CIF blocks have been arranged between files. The intent is that a site that archives pdCIFs will construct an index to `_pd_block_id` names that can be used to resolve block ID references.

The definition for `_pd_block_id` gives a procedure for creating a `_pd_block_id` name that is extremely unlikely to be duplicated. Other mechanisms for creating unique names can also be used: for example, using a web page name (URL) could be appropriate if care is taken never to reuse the URL.

The need for the block ID/block pointer mechanism is demonstrated by the following example. Consider a case where a neutron powder diffraction data set and an X-ray powder diffraction data set have been used together to determine a single structural model for a single crystalline phase. CIF does not allow the two data sets to be placed in a single block, since this would require two independent loops of observations where each loop uses some of the same data names. One can create a CIF with two blocks and include the structural model in the block that contains either of the two data sets. However, if this is done, a logical link is needed between the two blocks to make it clear that the structural model was derived from both data sets. It is better practice to place the

references