International Tables for Crystallography (2006). Vol. G, Section 3.2.2, pp. 92–98.

3.2. Classification and use of core data

BY S. R. HALL, P. M. D. FITZGERALD AND B. MCMAHON

3.2.1. Introduction

This chapter is concerned with the classification and organization of data items defined in the core CIF dictionary (Chapter 4.1). The core dictionary, as its name suggests, is central to the definition of data items found in most CIFs. It defines the measured and derived items common to most crystallographic experiments, analyses and publications, and, in particular, those items characterizing a classical single-crystal X-ray diffraction determination of a small-molecule or inorganic structure. As the nature of crystallographic studies evolves, so do the data items needed to describe them. New data names are introduced as needed to describe new techniques or technologies or simply to provide more details of subjects already covered. In addition, the developers of specialist dictionaries may find that some of the items they define have a wider application and propose that these items be added to the core dictionary instead.

Core data items are defined with two formalisms. The core dictionary, as presented in Chapter 4.1, defines core data items exclusively using the data definition language DDL1 (described in Chapter 2.5). However, core data items are also embedded within the macromolecular CIF dictionary presented in Chapter 4.5 using the data definition language DDL2 (described in Chapter 2.6). Because the revision cycles of the core and mmCIF dictionaries are not synchronized, at any one time the mmCIF dictionary may not include the complete set of data items in the current core dictionary. The mmCIF dictionary described in this volume includes the full content of core CIF dictionary version 2.3.1, also described in this volume.

The discussion in this chapter will concentrate on the current DDL1 version of the core dictionary (version 2.3, released on 4 October 2003 and reissued with minor amendments as version 2.3.1 in this volume). There will be some discussion of the more formal approach to the classification of data items that DDL2 permits.

In accordance with the scheme given in Table 3.1.10.1, groups of categories of data items in the core dictionary will be classified under the headings *Experimental measurements* (Section 3.2.2), *Analysis* (Section 3.2.3), *Atomicity, chemistry and structure* (Section 3.2.4), *Publication* (Section 3.2.5) and *File metadata* (Section 3.2.6). To help the reader relate the thematic order of the discussion of these categories to the alphabetic layout of the dictionary, the category structure of the core dictionary is summarized in Table 3.2.1.1 and is listed in full in Appendix 3.2.1. The appendix also lists for each category the section of this chapter in which the category is described.

The data items contained within each category are listed in the detailed commentary below. Where relevant, the data item or items that represent a unique identifier for a looped list ('category keys') are listed first and are marked by a bullet (•). Note that category keys are defined more formally in the mmCIF dictionary (see Chapter 2.6 and the discussion of categories in Section 3.1.6.4). The remaining data items in each category are listed alphabetically.

Table 3.2.1.1. *Category groups defined in the core CIF dictionary*

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

Section	Category group	Subject covered
(*a*) Experimental measurements		
3.2.2.1	CELL	Unit cell
3.2.2.2	DIFFRN	Diffraction experiment
3.2.2.3	EXPTL	Experimental conditions
(*b*) Analysis		
3.2.3.1	REFINE	Refinement procedures
3.2.3.2	REFLN	Reflection measurements
(*c*) Atomicity, chemistry and structure		
3.2.4.1	ATOM	Atom sites
3.2.4.2	CHEMICAL	Chemical properties and nomenclature
3.2.4.3	GEOM	Geometry of atom sites
3.2.4.4	SYMMETRY	Symmetry information
3.2.4.5	VALENCE	Bond-valence information
(*d*) Publication		
3.2.5.1	CITATION	Bibliographic references
3.2.5.2	COMPUTING	Computational details of the experiment
3.2.5.3	DATABASE	Database information
3.2.5.4	JOURNAL	Journal housekeeping
3.2.5.5	PUBL	Contents of a published article
(*e*) File metadata		
3.2.6	AUDIT	Dictionary maintenance and identification

3.2.2. Experimental measurements

Crystallographic archive files predating CIF were often constructed to serve the purposes of a particular software program or suite and stored the data generated by an experiment without providing a full record of the conditions under which the data were obtained. This is not unique to crystallography: many data formats make no provision for the metadata – information about the procedures for gathering and analysing data – that give context and in many cases significance to the numeric values. A specific goal of the design of CIF was to treat such supporting information as essential elements of the whole collection of information relating to a structure determination, rather than as optional and poorly defined metadata. There are therefore many categories in the core dictionary that relate to experimental conditions and apparatus, and these categories are discussed in this section. They include the categories in the DIFFRN group describing the traditional crystallographic diffraction experiment (typically a single-crystal laboratory-based X-ray determination, but increasingly including synchrotron experiments and experiments using other radiation types). There are also categories that describe and characterize the crystal used in the experiment and those that characterize the unit cell, since the experimental determination of the cell parameters is an essential part of the full structure-determination experiment.

Affiliations: SYDNEY R. HALL, School of Biomedical and Chemical Sciences, University of Western Australia, Crawley, 6009, Australia; PAULA M. D. FITZGERALD, Merck Research Laboratories, Rahway, New Jersey, USA; BRIAN MCMAHON, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

3.2.2.1. Crystal cell parameters and measurement conditions

The categories describing the crystal unit cell and its determination are as follows:

CELL group
- CELL
- CELL_MEASUREMENT_REFLN

The data items in these categories are as follows:

(*a*) CELL
- _cell_angle_alpha
- _cell_angle_beta
- _cell_angle_gamma
- _cell_formula_units_Z
- _cell_length_a
- _cell_length_b
- _cell_length_c
- _cell_measurement_pressure
- _cell_measurement_radiation
- _cell_measurement_reflns_used
- _cell_measurement_temperature
- _cell_measurement_theta_max
- _cell_measurement_theta_min
- _cell_measurement_wavelength
- _cell_reciprocal_angle_alpha
- _cell_reciprocal_angle_beta
- _cell_reciprocal_angle_gamma
- _cell_reciprocal_length_a
- _cell_reciprocal_length_b
- _cell_reciprocal_length_c
- _cell_special_details
- _cell_volume

(*b*) CELL_MEASUREMENT_REFLN
- • _cell_measurement_refln_index_h
- • _cell_measurement_refln_index_k
- • _cell_measurement_refln_index_l
- _cell_measurement_refln_theta

The bullet (•) *indicates a category key.*

The CELL category includes two groups of data names: those characterizing a crystal unit cell, and those describing the experimental conditions relating to the unit-cell determination. It is a feature of the formal definition of the *category* classification unit in CIF dictionaries that these may be classed within the same category, whereas the Miller indices of the reflections used in the measurement of the unit cell belong to a different category. An argument could be made for dividing the CELL category into two categories to reflect the division drawn above between the cell parameters and their determination. However, the CIF dictionaries have been designed to have as few separate categories as possible, subject to the constraint that data items that are looped together in the same list must belong to the same category.

The individual dictionary definitions of the data items in this category are unambiguous, with the possible exception of _cell_formula_units_Z, which records the number of complete *chemical* formula units present in the unit cell, and *not* the number of repetitions of the asymmetric unit. In some instances the value of Z could be less than the number of repetitions of the asymmetric unit, such as when an internally symmetric molecular unit is positioned on a symmetry element and spans multiple asymmetric units. Of course, Z can be greater than the number of repetitions of the asymmetric unit (*i.e.* $Z' > 1$).

Note that the value associated with the data item _cell_volume is not independent, but can be derived from the other cell parameters. Within the core dictionary there are many cases of derivable items, both because they have traditionally been reported separately and because the presence of redundant information allows cross checking of the internal consistency of the data set.

Data items in the CELL_MEASUREMENT_REFLN category record details about the reflections used to determine the crystallographic cell parameters. The key items in this list are marked with bullets; all three of the h, k, l values are needed to identify a reflection.

3.2.2.2. Data collection

The categories describing data collection are as follows:

DIFFRN group

General description (§3.2.2.2.1)
- DIFFRN

Apparatus and instrumentation before the crystal (§3.2.2.2.2)
- DIFFRN_ATTENUATOR
- DIFFRN_RADIATION
- DIFFRN_RADIATION_WAVELENGTH
- DIFFRN_SOURCE

Apparatus and instrumentation at the crystal (§3.2.2.2.3)
- DIFFRN_MEASUREMENT
- DIFFRN_ORIENT_MATRIX
- DIFFRN_ORIENT_REFLN

Apparatus and instrumentation after the crystal (§3.2.2.2.4)
- DIFFRN_DETECTOR

Intensity measurements (§3.2.2.2.5)
- DIFFRN_REFLN
- DIFFRN_REFLNS
- DIFFRN_REFLNS_CLASS
- DIFFRN_SCALE_GROUP
- DIFFRN_STANDARD_REFLN
- DIFFRN_STANDARDS

The category group related to the diffraction experiment is broad and includes details of the apparatus as well as the measurements. The individual categories are grouped together according to location within the experimental setup (see Fig. 3.2.2.1) or the measurement of intensities.

3.2.2.2.1. *General description*

The data items in this category are as follows:

DIFFRN
- _diffrn_ambient_environment
- _diffrn_ambient_pressure
- _diffrn_ambient_pressure_gt
- _diffrn_ambient_pressure_lt
- _diffrn_ambient_temperature
- _diffrn_ambient_temperature_gt
- _diffrn_ambient_temperature_lt
- _diffrn_crystal_treatment
- _diffrn_measured_fraction_theta_full
- _diffrn_measured_fraction_theta_max
- _diffrn_special_details
- _diffrn_symmetry_description

These data items give an overview of the diffraction experiment. They are intended to be independent of the instrument, techniques or methodology of the experiment.

The items describing the ambient environmental conditions are reasonably self-explanatory. They are often absent from a CIF, because an author has not thought it necessary to provide

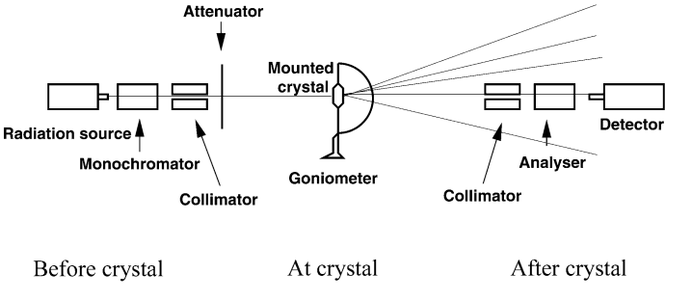

Fig. 3.2.2.1. Scheme of a diffraction experiment.

Example 3.2.2.1. *Attenuation of reflection intensities indicated by reference to attenuator scaling factors.*

```
loop_
     _diffrn_attenuator_code
     _diffrn_attenuator_material
     _diffrn_attenuator_scale
          1   Zr   16.976

loop_
     _diffrn_refln_index_h
     _diffrn_refln_index_k
     _diffrn_refln_index_l
     _diffrn_refln_attenuator_code
     _diffrn_refln_angle_chi
     _diffrn_refln_scan_rate
     _diffrn_refln_counts_bg_1
     _diffrn_refln_counts_total
     _diffrn_refln_counts_bg_2
     _diffrn_refln_scan_width
     _diffrn_refln_elapsed_time

0  0 -16  .   0.0 4.12  28 127  36 1.516 19.43
3  4  -4  1   0.0 1.03  69 459  73 1.516 2082.58
```

information for experiments conducted under 'normal' conditions of room temperature and pressure, and in a standard atmosphere. However, 'normal room temperature' may span a range of many degrees Kelvin and might have a non-negligible effect upon cell dimension measurements, so the temperature should be given. As there is significant variability in the ambient temperature at which laboratory experiments may be carried out, it is not appropriate to assign a default value for **_diffrn_ambient_temperature**, since any numeric value chosen as a default could be misconstrued as an experimentally determined value. If the ambient temperature has not been measured, an author may supply a best estimate of the ambient temperature with a suitable standard uncertainty. Alternatively, known upper and lower limits for the temperature may be given using **_diffrn_ambient_temperature_lt** and ***_gt**. The same considerations hold true for ambient pressure.

The default for **_diffrn_ambient_environment** may be understood as 'air', although formally it is impossible in the dictionary to specify a default for a free-text field.

The **_diffrn_measured_fraction_theta_*** items are provided in this category as an indication of the completeness of a set of reflection measurements. They are not as general as the other items in this category, as they apply only to monochromatic X-ray diffraction experiments, and they do not reflect the way macromolecular crystallographers tend to analyse the completeness of a data set as a function of resolution. When used, they must be accompanied by the value of the monochromatic radiation wavelength **_diffrn_radiation_wavelength** and relate to the maximum θ angle for which the measured reflection count is considered as complete (**_diffrn_reflns_theta_full**).

The other textual data items are provided for comment on other aspects of the handling of the crystal prior to the intensity measurement (**_diffrn_crystal_treatment**), observations on the diffraction point symmetry, systematic absences and inferred space group or superspace group relationships (**_diffrn_symmetry_description**) and any other comment on the intensity measurement process as a whole that cannot be accommodated elsewhere (**_diffrn_special_details**).

3.2.2.2.2. *Apparatus and instrumentation before the crystal*

The data items in these categories are as follows:

(*a*) DIFFRN_ATTENUATOR
- **_diffrn_attenuator_code**

_diffrn_attenuator_material
_diffrn_attenuator_scale

(*b*) DIFFRN_RADIATION
_diffrn_radiation_collimation
_diffrn_radiation_filter_edge
_diffrn_radiation_inhomogeneity
_diffrn_radiation_monochromator
_diffrn_radiation_polarisn_norm
_diffrn_radiation_polarisn_ratio
_diffrn_radiation_probe
_diffrn_radiation_type
_diffrn_radiation_xray_symbol

(*c*) DIFFRN_RADIATION_WAVELENGTH
- **_diffrn_radiation_wavelength_id**

_diffrn_radiation_wavelength
_diffrn_radiation_wavelength_wt

(*d*) DIFFRN_SOURCE
† **_diffrn_radiation_source**
_diffrn_source
_diffrn_source_current
_diffrn_source_details
_diffrn_source_power
_diffrn_source_size
_diffrn_source_take-off_angle
_diffrn_source_target
_diffrn_source_type
_diffrn_source_voltage

The bullet (•) *indicates a category key. The dagger* (†) *indicates a deprecated item, which should not be used in the creation of new CIFs.*

Attenuator properties are described by data items in the DIFFRN_ATTENUATOR category. Where an attenuator is used to reduce the intensity of an X-ray beam, this category may be used to describe the attenuator and its scaling factor. Details of multiple attenuator settings or materials can be included and each is identified by a code. A matching code value (**_diffrn_refln_attenuator_code**) appears in the list of intensities against each reflection that must be scaled by the appropriate attenuation factor. In Example 3.2.2.1, the intensity of the second reflection has been reduced using a zirconium attenuator and must be multiplied by 16.976 to place it on the same scale as the first (and other unattenuated intensities).

The DIFFRN_RADIATION category describes the radiation used in the diffraction experiment and its experimental handling by collimation and monochromatization before it interacts with the sample. [Post-sample treatment of the radiation beam after diffraction (including passage through any analyser or collimator) is described by data items in the complementary DIFFRN_DETECTOR category.] Many of the data items in this category are descriptive. Additional information about the generation of the radiation is also found in the DIFFRN_SOURCE category.

The use of **_diffrn_radiation_probe** is strongly recommended as an unambiguous indicator of the probing radiation or particle type (its permitted values are `x-ray`, `neutron`, `electron` and `gamma`). The similar-sounding data name **_diffrn_radiation_type** allows for a more detailed description of the radiation type, such as `white-beam` or (using the CIF code for the Greek character α, \a) `'Cu K\a'` for copper $K\alpha$ radiation. In the case of monochromatic (or near-monochromatic) X-radiation, a better representation is given by the use of **_diffrn_radiation_xray_symbol**, which can have one of a limited number of values expressing the X-ray wavelength according to IUPAC conventions (*e.g.* K-L_3, corresponding to the older Siegbahn notation $K\alpha_1$). If this data item is used, the element used as the X-ray generator target must also be specified using the data item **_diffrn_source_target**. Software for reading CIFs should be aware of these two alternative representations.

If the radiation beam is monochromatic, the wavelength can be provided using **_diffrn_radiation_wavelength**.

For a polychromatic beam, the other data items in the DIFFRN_RADIATION_WAVELENGTH category allow different wavelength components and an associated weighting factor for each component to be listed. In the list of experimental intensity measurements from a polychromatic beam (the DIFFRN_REFLN category, discussed below), each reflection has an associated **_diffrn_refln_wavelength_id** that must match the corresponding **_diffrn_radiation_wavelength_id** in this list.

The DIFFRN_SOURCE category specifies the characteristics of the radiation source in the experiment and is closely related to the DIFFRN_RADIATION category, which is concerned with the handling of the radiation beam before it reaches the specimen. (The now-deprecated data name **_diffrn_radiation_source** shows that there was no formal separation of the descriptions of the radiation generator and the radiation in the first release of the core dictionary.)

The general class of radiation is specified by the data name **_diffrn_source**, which is a free-text field. Typical entries would be 'sealed X-ray tube', 'nuclear reactor', 'synchrotron', 'spallation source', 'rotating-anode X-ray tube' or 'electron microscope'. It is clear that the category could describe non-X-ray experiments, but several of the data names within the category (*e.g.* **_diffrn_source_target**) have meanings that are specific to an X-ray experiment. New data names might be introduced if experiments using other radiation types become more common. For now, details that a user wishes to record that are not properly described by the existing data names may be stored in the **_diffrn_source_details** field.

3.2.2.2.3. *Apparatus and instrumentation at the crystal*

The data items in these categories are as follows:

(*a*) DIFFRN_MEASUREMENT

_diffrn_measurement_details
_diffrn_measurement_device
_diffrn_measurement_device_details
_diffrn_measurement_device_type
_diffrn_measurement_method
_diffrn_measurement_specimen_support

(*b*) DIFFRN_ORIENT_MATRIX

_diffrn_orient_matrix_type
_diffrn_orient_matrix_UB_11
_diffrn_orient_matrix_UB_12
_diffrn_orient_matrix_UB_13
_diffrn_orient_matrix_UB_21
_diffrn_orient_matrix_UB_22
_diffrn_orient_matrix_UB_23
_diffrn_orient_matrix_UB_31
_diffrn_orient_matrix_UB_32
_diffrn_orient_matrix_UB_33

(*c*) DIFFRN_ORIENT_REFLN

• **_diffrn_orient_refln_index_h**
• **_diffrn_orient_refln_index_k**
• **_diffrn_orient_refln_index_l**
_diffrn_orient_refln_angle_chi
_diffrn_orient_refln_angle_kappa
_diffrn_orient_refln_angle_omega
_diffrn_orient_refln_angle_phi
_diffrn_orient_refln_angle_psi
_diffrn_orient_refln_angle_theta

The bullet (•) *indicates a category key. Where multiple items within a category are marked with a bullet, they must be taken together to form a compound key.*

The DIFFRN_MEASUREMENT category currently concerns specifically the mounting of the crystal and the details of the goniometer or other device on which it is mounted, with the exception of **_diffrn_measurement_method**, which is defined simply as the 'method used to measure intensities'. In practice, for a typical single-crystal diffractometer setup this field is generally used to specify the scan type, as in Example 3.2.2.2, where the CIF code for the Greek character θ, \q, is used to indicate $\theta/2\theta$ scans.

Example 3.2.2.2. *An indication of the scan type of a diffractometer-based experiment.*

```
_diffrn_measurement_method
    'profile data from \q/2\q scans'
```

The orientation matrix gives the transformation between coordinates in a crystal-centric reference frame and those referred to the diffractometer axes. The data items defined in the DIFFRN_ORIENT_MATRIX category can be used to store the values in the matrix as recorded on an individual diffractometer and a reference to the convention used (in **_diffrn_orient_matrix_type**). However, the reference is not by itself sufficient to understand the transformation without additional external knowledge of the convention. Authors are encouraged to provide a full description of the convention in the text field **_diffrn_orient_matrix_type**.

The terminology UB refers to the conventional designation of the matrix relating reciprocal space and the reference frame of a diffractometer, calculated as the product of the orientation matrix **U** and the material matrix **B** by the method of Busing & Levy (1967).

The reflections used to determine the orientation matrix can be listed in the category DIFFRN_ORIENT_REFLN. As discussed above, this list is useful for analysing the results on a diffractometer of known type, but is not useful if the convention for establishing the individual terms of the orientation matrix is not known.

3.2.2.2.4. *Apparatus and instrumentation after the crystal*

The data items in this category are as follows:

DIFFRN_DETECTOR

_diffrn_detector
_diffrn_detector_area_resol_mean
_diffrn_detector_details
_diffrn_detector_dtime
_diffrn_detector_type
† **_diffrn_radiation_detector**
† **_diffrn_radiation_detector_dtime**

The dagger (†) *indicates a deprecated item, which should not be used in the creation of new CIFs.*

The DIFFRN_DETECTOR category is intended to describe the detector used to measure the scattered radiation, including any analyser and post-sample collimation. There are not many data names in this category, as it is not often necessary to know a lot about the detector beyond its make, model or name if it is made by a well known manufacturer. A record of the detector deadtime (**_diffrn_detector_dtime**) and the resolution of an area detector (**_diffrn_detector_area_resol_mean**) are useful details worth recording explicitly; other unusual or noteworthy details may be recorded in **_diffrn_detector_details**.

The deprecated items (retained for compatibility with the original release version) have been replaced by **_diffrn_detector** and **_diffrn_detector_dtime** to produce names better matched to the formal category assignment.

3.2.2.2.5. *Intensity measurements*

The data items in these categories are as follows:

(*a*) DIFFRN_REFLN

• **_diffrn_refln_index_h**
• **_diffrn_refln_index_k**
• **_diffrn_refln_index_l**

```
  _diffrn_refln_angle_chi
  _diffrn_refln_angle_kappa
  _diffrn_refln_angle_omega
  _diffrn_refln_angle_phi
  _diffrn_refln_angle_psi
  _diffrn_refln_angle_theta
  _diffrn_refln_attenuator_code
        → _diffrn_attenuator_code
  _diffrn_refln_class_code
        → _diffrn_reflns_class_code
  _diffrn_refln_counts_bg_1
  _diffrn_refln_counts_bg_2
  _diffrn_refln_counts_net
  _diffrn_refln_counts_peak
  _diffrn_refln_counts_total
  _diffrn_refln_crystal_id
        → _exptl_crystal_id
  _diffrn_refln_detect_slit_horiz
  _diffrn_refln_detect_slit_vert
  _diffrn_refln_elapsed_time
  _diffrn_refln_intensity_net
† _diffrn_refln_intensity_sigma
  _diffrn_refln_intensity_u
  _diffrn_refln_scale_group_code
        → _diffrn_scale_group_code
  _diffrn_refln_scan_mode
  _diffrn_refln_scan_mode_backgd
  _diffrn_refln_scan_rate
  _diffrn_refln_scan_time_backgd
  _diffrn_refln_scan_width
  _diffrn_refln_sint/lambda
  _diffrn_refln_standard_code
        → _diffrn_standard_refln_code
  _diffrn_refln_wavelength
  _diffrn_refln_wavelength_id
        → _diffrn_radiation_wavelength_id
```

(*b*) DIFFRN_REFLNS

```
  _diffrn_reflns_av_R_equivalents
† _diffrn_reflns_av_sigmaI/netI
  _diffrn_reflns_av_unetI/netI
  _diffrn_reflns_limit_h_max
  _diffrn_reflns_limit_h_min
  _diffrn_reflns_limit_k_max
  _diffrn_reflns_limit_k_min
  _diffrn_reflns_limit_l_max
  _diffrn_reflns_limit_l_min
  _diffrn_reflns_number
  _diffrn_reflns_reduction_process
  _diffrn_reflns_resolution_full
  _diffrn_reflns_resolution_max
  _diffrn_reflns_theta_full
  _diffrn_reflns_theta_max
  _diffrn_reflns_theta_min
  _diffrn_reflns_transf_matrix_11
  _diffrn_reflns_transf_matrix_12
  _diffrn_reflns_transf_matrix_13
  _diffrn_reflns_transf_matrix_21
  _diffrn_reflns_transf_matrix_22
  _diffrn_reflns_transf_matrix_23
  _diffrn_reflns_transf_matrix_31
  _diffrn_reflns_transf_matrix_32
  _diffrn_reflns_transf_matrix_33
```

(*c*) DIFFRN_REFLNS_CLASS

```
• _diffrn_reflns_class_code
  _diffrn_reflns_class_av_R_eq
† _diffrn_reflns_class_av_sgI/I
  _diffrn_reflns_class_av_uI/I
  _diffrn_reflns_class_d_res_high
  _diffrn_reflns_class_d_res_low
  _diffrn_reflns_class_description
  _diffrn_reflns_class_number
```

(*d*) DIFFRN_SCALE_GROUP

```
• _diffrn_scale_group_code
  _diffrn_scale_group_I_net
```

(*e*) DIFFRN_STANDARD_REFLN

```
• _diffrn_standard_refln_index_h
• _diffrn_standard_refln_index_k
• _diffrn_standard_refln_index_l
  _diffrn_standard_refln_code
```

(*f*) DIFFRN_STANDARDS

```
  _diffrn_standards_decay_%
  _diffrn_standards_interval_count
  _diffrn_standards_interval_time
  _diffrn_standards_number
† _diffrn_standards_scale_sigma
  _diffrn_standards_scale_u
```

The bullet (•) indicates a category key. Where multiple items within a category are marked with a bullet, they must be taken together to form a compound key. The arrow (→) 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.

The DIFFRN_REFLN category describes the measured reflections in a diffraction experiment. Example 3.2.2.3 shows a listing from a CAD-4 single-crystal diffractometer.

Note that the data in this list refer to the raw measurements as acquired at the time of data collection. This is in contrast to the data in the REFLN list, which refer to the reflections after merging and scaling.

The meanings of most of the data names can be deduced by inspection of this example. Full definitions are given in the dictionary.

However, the category also contains a number of data items which are used to group blocks of reflections with additional properties described by data items in other categories. For example, a number of reflections in the list might share a common value of `_diffrn_refln_scale_group_code`; this value would link to a description in the DIFFRN_SCALE_GROUP category of the scaling factor that needs to be applied to this group of reflections to bring all intensities in the list on to a common scale. (For example, intensities might be obtained from individual films in a multi-film data set or from a number of separate crystals.)

Likewise, individual reflections might be marked to indicate that they were monitored as standards during the course of the

Example 3.2.2.3. *Listing of experimental diffraction intensities.*

```
loop_
     _diffrn_refln_index_h
     _diffrn_refln_index_k
     _diffrn_refln_index_l
     _diffrn_refln_angle_chi
     _diffrn_refln_scan_rate
     _diffrn_refln_counts_bg_1
     _diffrn_refln_counts_total
     _diffrn_refln_counts_bg_2
     _diffrn_refln_angle_theta
     _diffrn_refln_angle_phi
     _diffrn_refln_angle_omega
     _diffrn_refln_angle_kappa
     _diffrn_refln_scan_width
     _diffrn_refln_elapsed_time
0  0 -16  0. 4.12  28 127  36  33.157  -75.846
          16.404   50.170  1.516 19.43
0  0 -15  0. 4.12  38 143  28  30.847  -75.846
          14.094   50.170  1.516 19.82
0  0 -14  0. 1.03 142 742 130  28.592  -75.846
          11.839   50.170  1.516 21.32
0  0 -13  0. 4.12  26 120  37  26.384  -75.846
           9.631   50.170  1.450 21.68
0  0 -12  0. 0.97 129 618 153  24.218  -75.846
           7.464   50.170  1.450 23.20
0  0 -11  0. 4.12  33 107  38  22.087  -75.846
           5.334   50.170  1.384 23.55
0  0 -10  0. 4.12  37 146  33  19.989  -75.846
           3.235   50.170  1.384 23.90
   # - - - - abbreviated - - - -
3  4  -4  0. 1.03  69 459  73  30.726  -53.744
          46.543  -47.552  1.516 2082.58
3  4  -5  0. 1.03  91 465  75  31.407  -54.811
          45.519  -42.705  1.516 2084.07
3 14  -6  0. 1.03  84 560  79  32.228  -55.841
          44.745  -38.092  1.516 2085.57
   # - - - - abbreviated - - - -
```

experiment, using the data name **_diffrn_refln_standard_code**. These standard reflections may be listed separately in the DIFFRN_STANDARD_REFLN category, in which case they are labelled by **_diffrn_standard_refln_code**, which must have values matching those assigned in the main list of intensities.

Apart from these specific classes of reflections, the intensity data may be binned according to different criteria (*e.g.* for modulated structures the intensities are often partitioned into classes with the same value of $m = \sum |m_i|$, where the m_i are the integer coefficients indexing diffraction vectors in an n-dimensional representation). The data name **_diffrn_refln_class_code** is provided as a link to the different classes of reflections defined in the DIFFRN_REFLNS_CLASS category.

The DIFFRN_REFLNS category describes collective properties of the set of experimental intensity measurements and follows the convention (common elsewhere in the dictionary) of having a name very similar to the related DIFFRN_REFLN category, but using a plural form of the relevant term in the composite name. While the individual DIFFRN_REFLN entries appear in a looped list, the items in the DIFFRN_REFLNS category are not looped.

This category describes properties of the *complete* measurement set; descriptions of specific portions of the complete set are handled by the DIFFRN_REFLNS_CLASS category.

Several of the items that appear in this category can be derived from the contents of the DIFFRN_REFLN lists, but it is often convenient to list them separately for ease of access and as a consistency check.

Note the definition of **_diffrn_reflns_number** as the total number of measured intensities *excluding* those classed as 'systematically absent' (reflections whose intensities are null as a consequence of crystallographic symmetry). There is no data item to specifically flag systematic absences (although one could assign a distinct **_diffrn_refln_class_code** value and define the relevant DIFFRN_REFLNS_CLASS). Because the measured diffraction data may (and often do) include reduced measurements and symmetry-equivalent reflection intensities, there is no formal way to check the value of **_diffrn_reflns_number** with dictionary-driven validation software. (Note that systematic absences *are* flagged in the structure-factor listing of the REFLN category.)

The data items in the DIFFRN_REFLNS_CLASS category record details about classes of reflections measured in the diffraction experiment. The user is free to assign classes according to arbitrary criteria; two specific cases, the marking of standard reflections and the clustering of intensities that need to be scaled by a common factor, have their own specific data items and associated categories, as discussed above. The example given in the dictionary (Example 3.2.2.4) describes a one-dimensional incommensurately modulated structure, where each reflection class is defined by the number $m = \sum |m_i|$, where the m_i are the integer coefficients that, in addition to h, k, l, index the corresponding diffraction vector in the basis defined for the reciprocal lattice.

The DIFFRN_SCALE_GROUP category records scaling factors which must be applied to specific intensities in the DIFFRN_REFLN list to bring all the measurements on to a common scale (Example 3.2.2.5). The scale factor **_diffrn_scale_group_I_net** is the factor by which the relevant net values in the intensities list must be multiplied. The intensities to which it must be applied are those in the intensities list marked with a **_diffrn_refln_scale_group_code** that matches the corresponding **_diffrn_scale_group_code** in this category.

The DIFFRN_STANDARD_REFLN category allows a separate tabulation of the reflections used as standards. Note that the actual *measurements* on these reflections are stored alongside all the other measurements in the DIFFRN_REFLN list. The results of the analysis of the standard reflections are described by the DIFFRN_STANDARDS category.

The DIFFRN_STANDARDS category describes the interval between measurements of the standard reflections and their overall intensity change (usually a decay, so that the relevant data name is **_diffrn_standards_decay_%**; this data item has a negative value if the final measured intensities are greater than the initial ones). The items assume a constant time interval (or number of counts) between the measurement of each standard and a single global value for the overall intensity change. If required, detailed tracking of the intensity change of individual standard reflections can be extracted from the DIFFRN_REFLN list provided the elapsed time at each measurement has been recorded (**_diffrn_refln_elapsed_time**).

Example 3.2.2.4. *Use of the DIFFRN_REFLNS_CLASS category to specify the main and satellite reflections collected for a modulated incommensurate structure*

```
loop_
     _diffrn_reflns_class_number
     _diffrn_reflns_class_d_res_high
     _diffrn_reflns_class_d_res_low
     _diffrn_reflns_class_av_R_eq
     _diffrn_reflns_class_code
     _diffrn_reflns_class_description
       1580 0.551 6.136 0.015 'Main'
          'm=0; main reflections'
       1045 0.551 6.136 0.010 'Sat1'
          'm=1; first-order satellites'
```

Example 3.2.2.5. *Scaling factors for reflections listed by group.*

```
loop_
     _diffrn_scale_group_code
     _diffrn_scale_group_I_net
      1    .86473
      2   1.0654
```

3.2.2.3. Experimental measurements on the crystal

The categories describing experimental conditions are as follows:

EXPTL group
- EXPTL
- EXPTL_CRYSTAL
- EXPTL_CRYSTAL_FACE

The data items in these categories are as follows:

(*a*) EXPTL

_exptl_absorpt_coefficient_mu
_exptl_absorpt_correction_T_max
_exptl_absorpt_correction_T_min
_exptl_absorpt_correction_type
_exptl_absorpt_process_details
_exptl_crystals_number
_exptl_special_details

(*b*) EXPTL_CRYSTAL

- **_exptl_crystal_id**

_exptl_crystal_colour
_exptl_crystal_colour_lustre
_exptl_crystal_colour_modifier
_exptl_crystal_colour_primary
_exptl_crystal_density_diffrn
_exptl_crystal_density_meas
_exptl_crystal_density_meas_gt
_exptl_crystal_density_meas_lt
_exptl_crystal_density_meas_temp
_exptl_crystal_density_meas_temp_gt
_exptl_crystal_density_meas_temp_lt
_exptl_crystal_density_method

```
_exptl_crystal_description
_exptl_crystal_F_000
_exptl_crystal_preparation
_exptl_crystal_pressure_history
_exptl_crystal_recrystallization_method
_exptl_crystal_size_length
_exptl_crystal_size_max
_exptl_crystal_size_mid
_exptl_crystal_size_min
_exptl_crystal_size_rad
_exptl_crystal_thermal_history
```

(*c*) EXPTL_CRYSTAL_FACE

```
• _exptl_crystal_face_index_h
• _exptl_crystal_face_index_k
• _exptl_crystal_face_index_l
  _exptl_crystal_face_diffr_chi
  _exptl_crystal_face_diffr_kappa
  _exptl_crystal_face_diffr_phi
  _exptl_crystal_face_diffr_psi
  _exptl_crystal_face_perp_dist
```

The bullet (•) indicates a category key. Where multiple items within a category are marked by a bullet, they must be taken together to form a compound key.

The EXPTL category is rather broadly named, but in practice is used to record details about any absorption correction applied and, using **_exptl_special_details**, any other details of the experimental work prior to intensity measurement not specifically described by other data items (*e.g.* **_exptl_crystal_preparation**).

The data items in the EXPTL_CRYSTAL category are designed to record details of experimental measurements on the crystal or crystals used. Since it is usually the case that just one crystal is used throughout the experiment, the category is presented as if it comprises non-looped data names. However, details of a number of crystals may be looped together, in which case **_exptl_crystal_id** is used to identify the different crystals and acts as the category key.

When different crystals are used to collect diffraction intensities, it is likely that the intensities collected from each crystal would need to be scaled by different factors, as recorded by the DIFFRN_SCALE_GROUP category and the **_diffrn_refln_scale_group_code** used for each individual reflection. In these circumstances, it would be good practice to use matching values of **_diffrn_refln_scale_group_code** and **_exptl_crystal_id**, although this is not mandatory.

Note that the $F(000)$ value, which is often calculated as the integer number of electrons in the crystal unit cell, may contain dispersion contributions and is more properly calculated as

$$F(000) = \left[\left(\sum f_r\right)^2 + \left(\sum f_i\right)^2\right]^{1/2},$$

where f_r and f_i are, respectively, the real and imaginary parts of the scattering factors at $\theta = 0$ and the sum is taken over each atom in the unit cell.

The crystal colour may be given as free text using the data item **_exptl_crystal_colour**. Alternatively, the standardized names developed by the International Centre for Diffraction Data to classify specimen colours may be constructed from the items **_exptl_crystal_colour_lustre**, **_*_modifier** and **_*_primary**, each of which has a restricted set of specific values.

The EXPTL_CRYSTAL_FACE category records details of the crystal faces. The faces are defined by Miller indices and their perpendicular distances from the centre of rotation of the crystal may be recorded in millimetres. Absolute orientations with respect to the goniometer angle settings may also be recorded. The category is currently constructed in a way that cannot distinguish between multiple crystals.

3.2.3. Analysis

The categories relevant to the structural analysis are as follows:

Refinement techniques and results (§3.2.3.1)

REFINE group

- REFINE
- REFINE_LS_CLASS

The reflections used in the refinement (§3.2.3.2)

REFLN group

- REFLN
- REFLNS
- REFLNS_CLASS
- REFLNS_SCALE
- REFLNS_SHELL

In the small-molecule and inorganic studies for which the core dictionary was designed, phasing and structure solution are almost routine, and the dictionary provides few specific fields for recording the details of the structure solution process: **_atom_sites_solution_primary**, **_atom_sites_solution_secondary** and **_atom_sites_solution_hydrogens** (Section 3.2.4.1.2); **_computing_structure_solution** (Section 3.2.5.2); and **_publ_section_exptl_solution** (Section 3.2.5.5). (In contrast, the macromolecular CIF includes extensive details of phasing.) Refinement, however, still allows for a wide range of techniques, practices and interpretation, and there are a large number of data names to allow a full account of the refinement strategy to be given. To complement this, several categories exist to provide a detailed listing and annotation of the structure factors and their treatment according to shells of resolution or other sorting criteria.

3.2.3.1. Structure refinement

The data items in these categories are as follows:

(*a*) REFINE

```
  _refine_diff_density_max
  _refine_diff_density_min
  _refine_diff_density_rms
  _refine_ls_abs_structure_details
  _refine_ls_abs_structure_Flack
  _refine_ls_abs_structure_Rogers
  _refine_ls_d_res_high
  _refine_ls_d_res_low
  _refine_ls_extinction_coef
  _refine_ls_extinction_expression
  _refine_ls_extinction_method
  _refine_ls_goodness_of_fit_all
  _refine_ls_goodness_of_fit_gt
† _refine_ls_goodness_of_fit_obs
  _refine_ls_goodness_of_fit_ref
  _refine_ls_hydrogen_treatment
  _refine_ls_matrix_type
  _refine_ls_number_constraints
  _refine_ls_number_parameters
  _refine_ls_number_reflns
  _refine_ls_number_restraints
  _refine_ls_R_factor_all
  _refine_ls_R_factor_gt
† _refine_ls_R_factor_obs
  _refine_ls_R_Fsqd_factor
  _refine_ls_R_I_factor
  _refine_ls_restrained_S_all
  _refine_ls_restrained_S_gt
† _refine_ls_restrained_S_obs
† _refine_ls_shift/esd_max
† _refine_ls_shift/esd_mean
  _refine_ls_shift/su_max
  _refine_ls_shift/su_max_lt
  _refine_ls_shift/su_mean
  _refine_ls_shift/su_mean_lt
  _refine_ls_structure_factor_coef
  _refine_ls_weighting_details
```


references