International Tables for Crystallography (2006). Vol. G, Chapter 3.2, pp. 92–116.

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

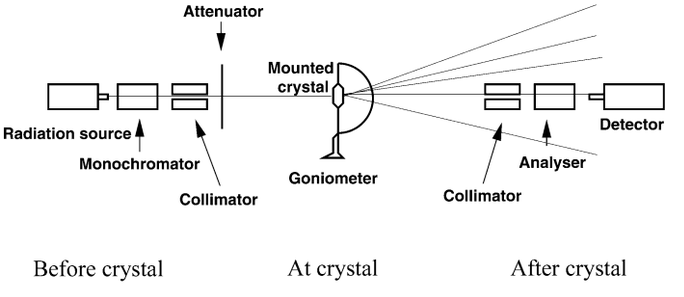

Fig. 3.2.2.1. Scheme of a diffraction experiment.

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
```

```
  _refine_ls_weighting_scheme
  _refine_ls_wR_factor_all
  _refine_ls_wR_factor_gt
† _refine_ls_wR_factor_obs
  _refine_ls_wR_factor_ref
  _refine_special_details
```

(*b*) REFINE_LS_CLASS

```
• _refine_ls_class_code
        → _reflns_class_code
  _refine_ls_class_d_res_high
  _refine_ls_class_d_res_low
  _refine_ls_class_R_factor_all
  _refine_ls_class_R_factor_gt
  _refine_ls_class_R_Fsqd_factor
  _refine_ls_class_R_I_factor
  _refine_ls_class_wR_factor_all
```

The bullet (•) indicates a category 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.

Example 3.2.3.1 shows how the data names in the REFINE category are used. Most of the dictionary entries are detailed and fully explanatory, so only a few points that might require special care are mentioned here.

Two groups of older data names have been superseded by new names that are functionally equivalent, but represent a more correct terminology. One group is of names that include the component '**_obs**' used to indicate 'observed' reflections; this has been replaced by the component '**_gt**' indicating that the measured values are greater than a threshold recorded elsewhere (as the value of **_reflns_threshold_expression**). The other group replaces the component '**_esd**' (for estimated standard deviation) with '**_su**' (for standard uncertainty).

A number of data names describe the extinction coefficient and the method used to determine it. Note that a default value (Zachariasen) is given in the dictionary for the method (**_refine_ls_extinction_method**); this *only* makes sense if this data item is missing from the data block but a value of **_refine_ls_extinction_coef** is present. This can complicate the design of software to read CIFs, which might assign to any missing data name a default value given by the dictionary.

Care is also needed with **_refine_ls_hydrogen_treatment**, which describes the treatment of hydrogen atoms in the refinement. Clearly, the data item only has meaning if there were hydrogen atoms in the model (although, since in this case the default value is undef for 'undefined', it could be argued that the default is appropriate even when hydrogen atoms were not included in the model).

The weighting scheme used in the refinement is described by the two data names **_refine_ls_weighting_scheme** and **_refine_ls_weighting_details**. The first of the two can take only one of the three values sigma (weights assigned based on measured standard uncertainties), unit (unit or no weights applied) or calc (calculated weights applied). The actual mathematical expression used in the weighting scheme should be stated in **_refine_ls_weighting_details**.

A wide variety of 'residual structure-factor difference measures', referred to as R factors, are used in crystallography as indicators of refinement quality. The core CIF dictionary contains definitions for the three most commonly used R factors. The 'conventional R factor' is defined as

$$R = \frac{\sum |F(\text{meas.}) - F(\text{calc.})|}{\sum |F(\text{meas.})|},$$

Example 3.2.3.1. *Summary of refinement results.*

```
_refine_special_details
                sfls:_F_calc_weight_full_matrix

_refine_ls_structure_factor_coef  F
_refine_ls_matrix_type            full
_refine_ls_weighting_scheme       calc
_refine_ls_weighting_details
                  'w=1/(u^2^(F)+0.0004F^2^)'
_refine_ls_hydrogen_treatment     refxyz
_refine_ls_extinction_method      Zachariasen
_refine_ls_extinction_coef        3514(42)
_refine_ls_extinction_expression
;Larson, A. C. (1970). "Crystallographic Computing",
 edited by F. R. Ahmed. Eq. (22) p. 292. Copenhagen:
 Munksgaard.
;
_refine_ls_abs_structure_details
;The absolute configuration was assigned to agree
 with that of its precursor l-leucine at the
 chiral centre C3.
;
_refine_ls_number_reflns          1408
_refine_ls_number_parameters      272
_refine_ls_number_restraints      0
_refine_ls_number_constraints     0
_refine_ls_R_factor_all           .038
_refine_ls_R_factor_gt            .034
_refine_ls_wR_factor_all          .044
_refine_ls_wR_factor_gt           .042
_refine_ls_goodness_of_fit_all   1.462
_refine_ls_goodness_of_fit_gt    1.515
_refine_ls_shift/su_max           .535
_refine_ls_shift/su_mean          .044
_refine_diff_density_min         -.108
_refine_diff_density_max          .131
```

where $F(\text{meas.})$ and $F(\text{calc.})$ are the measured and calculated structure factors, respectively. In the data item **_refine_ls_R_factor_all**, the sum used in the calculation is taken over all the reflections collected, whereas in the data item **_refine_ls_R_factor_gt**, the sum is taken over reflections with a value greater than the limit specified by **_refine_threshold_expression**. In both cases, the reflections included in the calculation may be limited to those between specified resolution limits.

This R factor is calculated from the F values, regardless of whether the structure-factor coefficient $|F|$, $|F|^2$ or I was actually used in the refinement, and is often taken as a convenient indicator of the relative quality of a structure determination. As most structure refinements used to be performed on $|F|$, it allows a structure determined today to be compared with an older study.

Many refinements are now carried out on $|F|^2$, although some may still use the absolute value of the structure factor $|F|$ or the net intensity I. The weighted residual factor wR and goodness of fit S for a refinement should be reported according to the coefficients actually used in the refinement. For example, the weighted residual over all reflections, **_refine_ls_wR_factor_all**, is defined as

$$wR = \left(\frac{\sum w[Y(\text{meas.}) - Y(\text{calc.})]^2}{\sum wY(\text{meas.})^2} \right)^{1/2},$$

where w represents the weights and Y represents the structure-factor coefficient, either $|F|$, $|F|^2$ or I as specified by **_refine_ls_structure_factor_coef**.

This distinction between the conventional R factor, which is invariably calculated using F values, and the wR and S factors also holds for similar expressions defined on subsets of the reflections, *e.g.* **_reflns_class_wR_factor_all**.

Note that data names are also provided for reporting *unweighted* residuals on $|F|^2$ or I, but these are rarely used in practice, with

Example 3.2.3.2. *Structure-factor listing.*

```
loop_
     _refln_index_h
     _refln_index_k
     _refln_index_l
     _refln_F_squared_calc
     _refln_F_squared_meas
     _refln_F_squared_sigma
     _refln_include_status
    2   0   0       85.57       58.90      1.45 o
    3   0   0    15718.18    15631.06     30.40 o
    4   0   0    55613.11    49840.09     61.86 o
    5   0   0      246.85      241.86     10.02 o
    6   0   0       82.16       69.97      1.93 o
    7   0   0     1133.62      947.79     11.78 o
    8   0   0     2558.04     2453.33     20.44 o
    9   0   0      283.88      393.66      7.79 o
    10  0   0      283.70      171.98      4.26 o
```

the exception of $R(I)$ in Rietveld refinements against powder data, where it is generally called the Bragg R factor, $R_{\rm Bragg}$ or R_B.

The data items in the REFINE_LS_CLASS category are similar to several in the general REFINE category, but correspond to values for separate reflection classes as described in the REFLNS_CLASS category. The data name **_refine_ls_class_code** identifies the individual classes through a direct match with a corresponding value of **_reflns_class_code**.

3.2.3.2. Reflection measurements

The categories describing the reflections used in the refinement are as follows:

REFLN group

Individual reflections (§3.2.3.2.1)

REFLN

Groups of reflections (§3.2.3.2.2)

REFLNS

REFLNS_CLASS

REFLNS_SCALE

REFLNS_SHELL

The main category in this group is REFLN, which stores the list of reflections used in the structure refinement process, their associated structure factors and information about how each reflection was handled. The distinction between the REFLN (singular) category and the REFLNS (plural) category parallels the distinction between the categories DIFFRN_REFLN and DIFFRN_REFLNS: data items in the REFLN category store information about individual reflections, while data items in the REFLNS category store information about the complete set of reflections, or about subsets of reflections selected by shells of resolution, scaling factors or other criteria.

3.2.3.2.1. *Individual reflections*

The data items in this category are as follows:

REFLN

- **_refln_index_h**
- **_refln_index_k**
- **_refln_index_l**

_refln_A_calc
_refln_A_meas
_refln_B_calc
_refln_B_meas
_refln_class_code
→ **_reflns_class_code**
_refln_crystal_id
→ **_exptl_crystal_id**
_refln_d_spacing
_refln_F_calc
_refln_F_meas
_refln_F_sigma
_refln_F_squared_calc
_refln_F_squared_meas
_refln_F_squared_sigma
_refln_include_status
_refln_intensity_calc
_refln_intensity_meas
_refln_intensity_sigma
_refln_mean_path_length_tbar
† **_refln_observed_status**
_refln_phase_calc
_refln_phase_meas
_refln_refinement_status
_refln_scale_group_code
→ **_reflns_scale_group_code**
_refln_sint/lambda
_refln_symmetry_epsilon
_refln_symmetry_multiplicity
_refln_wavelength
_refln_wavelength_id
→ **_diffrn_radiation_wavelength_id**

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.

Example 3.2.3.2 shows a typical structure-factor listing produced by a refinement program. This kind of structure-factor listing is suitable for deposition with a journal or a database. The Miller indices for each reflection are accompanied by the calculated and measured values of the quantity used in the refinement, and the standard uncertainty derived from the measurement. There is also an indication of whether each reflection was included in the refinement and in the calculation of R factors.

In this example, the squared structure factors $|F|^2$ are listed. When refinement is performed against the structure factors F or the intensities I, the data items **_refln_F_calc** or **_refln_intensity_calc** and the corresponding data names for the measured values and uncertainties should be used.

Individual calculated structure-factor components $A = |F|\cos\varphi$ and $B = |F|\sin\varphi$ may also be listed, along with the phase φ, using the data names **_refln_A_calc**, **_refln_B_calc** and **_refln_phase_calc**. Corresponding measured values have equivalent ***_meas** names.

The **_refln_include_status** flag is used to indicate whether reflections were used in the refinement and in the calculation of R factors, and if they were not used, to give the reason for exclusion of the reflection from the refinement. The flag o, which indicates that a reflection was used in the refinement, was originally chosen to indicate that the value of the reflection was higher than the limit specified by **_reflns_observed_criterion** and that the reflection was thus 'observed'. The data item **_reflns_observed_criterion** is now deprecated in favour of **_reflns_threshold_status**, and the value o is now taken to indicate not only that the reflection has an intensity suitable for inclusion in the refinement, but also that the reflection satisfies all other criteria used to select reflections for inclusion in the refinement.

Various other flags indicate reflections that were not included in the refinement. Reflections outside the range of d spacings bounded by the values **_refine_ls_d_res_high** and **_refine_ls_d_res_low** are flagged with h or l, respectively. Reflections within the resolution limits but below the intensity threshold are flagged with <. Systematically absent reflections are flagged with -. Sometimes a value can be identified as having a systematic error; these reflections can be flagged with x. However, great care must be taken in excluding reflections with apparently 'anomalous' structure factors (*i.e.* where the measured values are substantially different from the calculated ones), so as not to introduce bias into the refinement.

The flag **_refln_refinement_status** is used specifically to indicate whether a reflection was included in or excluded from

the refinement. Use of **_refln_include_status** to provide more information about each reflection is greatly preferred.

Other data names in this category allow the recording of specific information about each reflection, such as the symmetry reinforcement factor ε, the number of reflections symmetry-equivalent under the Laue symmetry, the d spacing, the mean path length through the crystal $\bar{t}$, the $(\sin\theta)/\lambda$ value and, in the case of Laue experiments, the mean wavelength of the radiation. (For polychromatic radiation, the wavelength information might instead be given by **_refln_wavelength_id**, which is a code identifying a matching entry in the DIFFRN_RADIATION category.)

Other codes provide links to identifiers in other categories. The **_refln_class_code** identifies a set of reflections binned as described by entries in the REFLNS_CLASS category. **_refln_scale_group_code** identifies groups of reflections to which the same structure-factor scaling has been applied.

Note that the values of the Miller indices in this list must correspond to the cell defined by the lengths and angles recorded in the CELL category; they may, however, be different from the Miller indices in the DIFFRN_REFLN list if a transformation of the original cell has taken place. In this case, the transformation matrix is given using the **_diffrn_reflns_transf_matrix_*** items.

The usual use of a CIF as an archive of a completed structure determination implies that the values given in the REFLN list are derived from the final cycle of refinement, but this is not a formal requirement. Care should be taken when preparing a CIF for archiving that the structural model corresponds to the refinement cycle summarized in the accompanying REFLN table, especially if the file is constructed from fragments output from different programs.

3.2.3.2.2. *Groups of reflections*

The data items in these categories are as follows:

(*a*) REFLNS

```
  _reflns_d_resolution_high
  _reflns_d_resolution_low
  _reflns_Friedel_coverage
  _reflns_limit_h_max
  _reflns_limit_h_min
  _reflns_limit_k_max
  _reflns_limit_k_min
  _reflns_limit_l_max
  _reflns_limit_l_min
  _reflns_number_gt
† _reflns_number_observed
  _reflns_number_total
† _reflns_observed_criterion
  _reflns_special_details
  _reflns_threshold_expression
```

(*b*) REFLNS_CLASS

```
• _reflns_class_code
  _reflns_class_d_res_high
  _reflns_class_d_res_low
  _reflns_class_description
  _reflns_class_number_gt
  _reflns_class_number_total
  _reflns_class_R_factor_all
  _reflns_class_R_factor_gt
  _reflns_class_R_Fsqd_factor
  _reflns_class_R_I_factor
  _reflns_class_wR_factor_all
```

(*c*) REFLNS_SCALE

```
• _reflns_scale_group_code
  _reflns_scale_meas_F
  _reflns_scale_meas_F_squared
  _reflns_scale_meas_intensity
```

(*d*) REFLNS_SHELL

```
• _reflns_shell_d_res_high
• _reflns_shell_d_res_low
† _reflns_shell_meanI_over_sigI_all
† _reflns_shell_meanI_over_sigI_gt
† _reflns_shell_meanI_over_sigI_obs
  _reflns_shell_meanI_over_uI_all
  _reflns_shell_meanI_over_uI_gt
  _reflns_shell_number_measured_all
  _reflns_shell_number_measured_gt
† _reflns_shell_number_measured_obs
  _reflns_shell_number_possible
  _reflns_shell_number_unique_all
  _reflns_shell_number_unique_gt
† _reflns_shell_number_unique_obs
  _reflns_shell_percent_possible_all
  _reflns_shell_percent_possible_gt
† _reflns_shell_percent_possible_obs
  _reflns_shell_Rmerge_F_all
  _reflns_shell_Rmerge_F_gt
† _reflns_shell_Rmerge_F_obs
  _reflns_shell_Rmerge_I_all
  _reflns_shell_Rmerge_I_gt
† _reflns_shell_Rmerge_I_obs
```

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

The data items in the REFLNS category describe properties or attributes of the complete set of reflections used in the structure refinement. Several are derivative and may be obtained from the information in the reflections list, but it is convenient to present them separately so that they do not need to be calculated again. They can also be used to check the consistency of the reflections list.

The **_reflns_limit_*** data items define the upper and lower bounds on the Miller indices and on the interplanar d spacings.

The **_reflns_threshold_expression** is a text field describing the criterion applied to mark individual reflections as 'significantly intense' (*i.e.* distinct from the background level). This is typically expressed as a multiple of the standard uncertainty on the quantity used in refinement, *e.g.* `I>2u(I)`.

The number of reflections with values higher than the threshold is reported in **_reflns_number_gt**. The total number of reflections measured is given by **_reflns_number_total**. Although the use of these data names appears to be obvious, different practices have been used in the past to report total numbers (*e.g.* by neglecting symmetry-equivalent reflections) and the definitions in the dictionary should be consulted. Both numbers may contain Friedel-equivalent reflections (those which are symmetry-equivalent under the Laue symmetry but inequivalent under the crystal class).

The proportion of Friedel-related reflections present is reported separately by **_reflns_Friedel_coverage**, defined as $(N_C - N_L)/N_L$, where N_C is the number of reflections obtained on averaging under the symmetry of the crystal class and N_L is the number obtained on averaging under the Laue class. The definition in the dictionary provides examples of how the value of this data name may be used as an indicator of the fraction of the available reciprocal space sampled in the diffraction experiment.

The deprecated data names **_reflns_observed_criterion** and **_reflns_number_observed** reflect the old use of 'observed' as a term describing significantly intense reflections. They should not be used in the creation of new CIFs, but are retained to ensure that the information can be extracted from old CIFs.

The free-text field **_reflns_special_details** can be used to discuss any aspects of the reflections list not covered by other data names. It is recommended that information about the averaging of symmetry-equivalent reflections (including Friedel pairs) should be given here.

Example 3.2.3.3. *Description of subsets of the reflection list.*

```
loop_
  _reflns_class_number_gt
  _reflns_class_code
  _reflns_class_description
      584    'Main'   'm=0; main reflections'
      226    'Sat1'   'm=1; first-order satellites'
      50     'Sat2'   'm=2; second-order satellites'
```

The REFLNS_CLASS category is used to summarize the properties of subsets of the reflection list. The data names are analogous to several in the REFLNS and REFINE categories, but are applied to individual classes of reflections labelled by **_reflns_class_code** and described by **_reflns_class_description** (see Example 3.2.3.3).

Individual reflections in the structure-factor listing can be recognized through the matching value of **_refln_class_code** as belonging to a particular class labelled by **_reflns_class_code**.

Although classes can be assigned according to arbitrary criteria, the specific case for which the REFLNS_CLASS category was designed was the partitioning of the reflection list into contributions from different components in incommensurately modulated structures. However, the formalism is general and other binning strategies can be described. Note, however, that the specific case of processing of reflections by shells of resolution (in macromolecular crystallography, for example) is handled explicitly by the REFLNS_SHELL category.

The category REFLNS_SCALE provides a listing of the scale factors applied to individual reflections sharing a common value of **_refln_scale_group_code**. Each value is indexed by the matching identifier **_reflns_scale_group_code** of this category.

The REFLNS_SHELL category describes the properties of separate resolution shells of reflections and is a special case of the binning of reflections into classes (compare REFLNS_CLASS above).

Each shell is defined by an upper and lower resolution limit (**_reflns_shell_d_res_high** and ***_low**), and for each shell there are data names for the number of reflections measured and exceeding a threshold of significance, for the percentage of geometrically possible reflections collected, and for the ratios of the mean intensities to their standard uncertainties.

$R_{\rm merge}$ values are also defined for each shell of resolution (both for all measured reflections and for significantly intense ones).

This category also contains a number of deprecated data names reflecting older terminology and notation. Such data names should not be used in creating new CIFs, but will need to be recognized by CIF-reading software in order to process old CIFs.

3.2.4. Atomicity, chemistry and structure

The core CIF dictionary provides many data names for describing the structural model.

The categories describing the atom sites handle these in a general way as sites of significant electron density which might be contributed to by more than one element species. The chemical identification of the compound under study, and where appropriate a model of the molecular connectivity and bonding, are handled separately by the chemistry-related categories. The geometry-related categories are purely derivative, given knowledge of the positions of the atom sites and the crystallographic symmetry; but as with other examples of derived data, they are given their own data names to provide convenient listings and to check the consistency of information provided by other categories. The symmetry-related data names in the core dictionary are restricted to those essential for the construction of a geometric model; Chapter 3.8 describes a symmetry extension dictionary suitable for a more complete description of crystal symmetry.

3.2.4.1. Atom sites

The categories describing atom sites are as follows:

ATOM group

Individual atom sites (§3.2.4.1.1)

ATOM_SITE

Collections of atom sites (§3.2.4.1.2)

ATOM_SITES

Atom types (§3.2.4.1.3)

ATOM_TYPE

These categories permit the traditional interpretation of regular concentrations of electron density in a crystalline lattice as atom sites containing one or more chemical elements, with complete or partial occupancy, and with a spatial distribution affected by thermal displacement or disorder.

Lists of atom-site coordinates and anisotropic displacement factors are covered by data items in the ATOM_SITE category. Identification of the chemical species occupying each site is handled by data items in the ATOM_TYPE category and data items in the ATOM_SITES category record collective information common to all sites.

While the ATOM_SITE category formally contains the data items describing both positions and atomic displacements, the anisotropic displacement parameters are often given in a separate looped list. In the version of the core dictionary embedded in the macromolecular CIF dictionary, which uses the DDL2 formalism, this is recognized by the creation of a separate, but overlapping, ATOM_SITE_ANISOTROP category.

3.2.4.1.1. *Individual atom sites*

The data items in this category are as follows:

ATOM_SITE

```
• _atom_site_label
  _atom_site_adp_type
  _atom_site_aniso_B_11
  _atom_site_aniso_B_12
  _atom_site_aniso_B_13
  _atom_site_aniso_B_22
  _atom_site_aniso_B_23
  _atom_site_aniso_B_33
  _atom_site_aniso_label
        → _atom_site_label
  _atom_site_aniso_ratio
  _atom_site_aniso_type_symbol
        → _atom_site_type_symbol
  _atom_site_aniso_U_11
  _atom_site_aniso_U_12
  _atom_site_aniso_U_13
  _atom_site_aniso_U_22
  _atom_site_aniso_U_23
  _atom_site_aniso_U_33
  _atom_site_attached_hydrogens
  _atom_site_B_equiv_geom_mean
  _atom_site_B_iso_or_equiv
  _atom_site_calc_attached_atom
  _atom_site_calc_flag
  _atom_site_Cartn_x
  _atom_site_Cartn_y
  _atom_site_Cartn_z
  _atom_site_chemical_conn_number
        → _chemical_conn_atom_number
  _atom_site_constraints
  _atom_site_description
  _atom_site_disorder_assembly
  _atom_site_disorder_group
  _atom_site_fract_x
  _atom_site_fract_y
  _atom_site_fract_z
  _atom_site_label_component_0
  _atom_site_label_component_1
```

```
  _atom_site_label_component_2
  _atom_site_label_component_3
  _atom_site_label_component_4
  _atom_site_label_component_5
  _atom_site_label_component_6
  _atom_site_occupancy
† _atom_site_refinement_flags
  _atom_site_refinement_flags_adp
  _atom_site_refinement_flags_occupancy
  _atom_site_refinement_flags_posn
  _atom_site_restraints
  _atom_site_symmetry_multiplicity
† _atom_site_thermal_displace_type
  _atom_site_type_symbol
        → _atom_type_symbol
  _atom_site_U_equiv_geom_mean
  _atom_site_U_iso_or_equiv
  _atom_site_Wyckoff_symbol
```

The bullet (•) indicates a category key. For this category an alternative category key can be formed by taking all the **_atom_site_label_component_*** *items together. Anisotropic displacement parameters may also be listed in a separate loop, for which* **_atom_site_aniso_label** *forms the 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.*

Data items in the ATOM_SITE category represent the positions of atom sites identified in the structural model, their spatial distribution defined by isotropic or anisotropic displacement parameters, details of restraints or constraints applied during the refinement, and the interpretation of their occupancy due to structural or compositional disorder.

Example 3.2.4.1 is a typical extract from a list of atom-site coordinates, with equivalent isotropic displacement values and refinement conditions. Each site is identified by **_atom_site_label**.

The coordinates are specified as fractional x, y, z values along the unit-cell axes. Coordinates may also be specified in ångström units along orthogonal Cartesian axes using the data names **_atom_site_Cartn_x**, **_atom_site_Cartn_y** and **_atom_site_Cartn_z**. The transformation matrix between Cartesian and fractional coordinates can be given in the ATOM_SITES category.

(Note that occupancy values are unaffected by symmetry. This is discussed later in connection with site multiplicity.)

_atom_site_U_iso_or_equiv records the isotropic atomic displacement value $U_{\rm iso}$ in the case of isotropic refinement. In the case of anisotropic refinement, **_atom_site_U_iso_or_equiv** records the equivalent isotropic value $U_{\rm eq}$, defined as

$$U_{\rm eq} = (1/3)\sum_i \left[\sum_j (U^{ij} a_i^* a_j^* a_i a_j)\right],$$

where a_i are the real-space cell lengths, a_j^* are the reciprocal-space cell lengths and U^{ij} are the anisotropic displacement parameters.

The data item **_atom_site_adp_type** identifies which value is given. An alternative equivalent isotropic displacement parameter **_atom_site_U_equiv_geom_mean** may be calculated as the geometric mean of the anisotropic parameters,

$$U_{\rm eq} = (U_i U_j U_k)^{1/3},$$

where the U_i are the principal components of the orthogonalized U^{ij}.

Data names also exist for the corresponding quantities calculated from B values, although the use of B values is discouraged by the IUCr Commission on Crystallographic Nomenclature.

For each site, **_atom_site_calc_flag** takes one of the following values: `d`, to indicate that the atom-site coordinates were determined from the diffraction intensities; `c` or `calc` to indicate that they were calculated from molecular geometry considerations; or `dum`, for a dummy site.

Example 3.2.4.1. *List of atom-site coordinates, equivalent isotropic U values and refinement conditions.*

```
loop_
     _atom_site_label
     _atom_site_fract_x
     _atom_site_fract_y
     _atom_site_fract_z
     _atom_site_U_iso_or_equiv
     _atom_site_adp_type
     _atom_site_calc_flag
     _atom_site_refinement_flags_posn
     _atom_site_occupancy
     _atom_site_disorder_assembly
     _atom_site_disorder_group
     _atom_site_type_symbol
   O1 0.5000 1.0000 0.8011(2) 0.0259(6)
      Uani d    S 1 . . O
   O2 0.33569(10) 0.98239(10) 0.88892(17) 0.0321(5)
      Uani d    . 1 . . O
   O3 0.20150(13) 0.92560(11) 0.8817(2) 0.0458(5)
      Uani d    . 1 . . O
   O4 0.35539(11) 0.81530(10) 0.96958(17) 0.0333(5)
      Uani d    . 1 . . O
   C1 0.43883(16) 0.95672(14) 0.7293(3) 0.0275(6)
      Uani d    . 1 . . C
   H1 0.4064 0.9930 0.6730 0.0493(19)
      Uiso calc R 1 . . H
   C2 0.37292(16) 0.92010(14) 0.8175(3) 0.0266(6)
      Uani d    . 1 . . C
   H2 0.3246 0.8945 0.7691 0.0493(19)
      Uiso calc R 1 . . H
   C3 0.41827(16) 0.85968(14) 0.8983(3) 0.0280(6)
      Uani d    . 1 . . C
   H3 0.4629 0.8850 0.9536 0.0493(19)
      Uiso calc R 1 . . H
```

Specific restraints or constraints applied to a site may be indicated by one or more of the **_atom_site_refinement_flags_*** items.

The data item **_atom_site_occupancy** defines the fraction of the atom type present at the site. Note that the same site may occur more than once in the list, identified by separate values of **_atom_site_label**. Such an arrangement would represent contributions from separate atom types (perhaps in modelling compositional disorder). The sum of occupancies of all atom types present at a single site may not significantly exceed 1.0 (unless it is a dummy site with no physical significance). Note that an atom of a given chemical species positioned on a special position (*e.g.* on a twofold axis) will in general be assigned a full occupancy value of 1.0. However, it will occur less often in the unit cell than an atom on a general position (in this example by a factor of 2). To account for this in structure-factor calculations it may be given a population value of 0.5 within the refinement program. A population adjustment of this kind is *not* implied in the assignment of a value to **_atom_site_occupancy**. The multiplicity of the site owing to the space-group symmetry can be recorded in **_atom_site_symmetry_multiplicity**.

The disorder-related data names in this example will be discussed below.

_atom_site_type_symbol is a code which must match an entry in the ATOM_TYPE category that supplies information about the elemental composition and scattering factors of the atom or atoms occupying the site. Note that it is quite legitimate to have an atom-type symbol such as 'Fe3+Ni2+', referring to a mixed-composition atom site. The effective physical properties of such a pseudo-atom should be given in full in the ATOM_TYPE category.

Example 3.2.4.2 demonstrates how the anisotropic displacement parameters are conventionally broken out into a separate list. When this is done, each atom site is identified by **_atom_site_aniso_label**, and this must of course match the value of **_atom_site_label** specifying the position of the site.

Example 3.2.4.2. *Separate list of anisotropic U values with* **_atom_site_aniso_label** *acting as the key that uniquely identifies table rows in this listing.*

```
loop_
     _atom_site_aniso_label
     _atom_site_aniso_U_11
     _atom_site_aniso_U_22
     _atom_site_aniso_U_33
     _atom_site_aniso_U_12
     _atom_site_aniso_U_13
     _atom_site_aniso_U_23
    O1 0.0256(9) 0.0275(9) 0.0433(12) 0.0020(8)
       0.0038(8) -0.0062(9)
    O2 0.0306(10) 0.0418(11) 0.0651(15) -0.0037(10)
       0.0079(11) -0.0049(11)
    O3 0.0374(10) 0.0309(10) 0.0315(11) -0.0030(9)
       0.0051(9) 0.0005(9)
    C1 0.0279(13) 0.0245(13) 0.0300(15) 0.0004(11)
       -0.0020(12) -0.0034(12)
    C2 0.0258(13) 0.0226(13) 0.0314(15) 0.0024(11)
       -0.0024(12) -0.0033(12)
    C3 0.0292(13) 0.0246(13) 0.0302(16) -0.0028(11)
       0.0031(12) 0.0005(12)
```

The data item **_atom_site_label** is normally used as the identifier of each individual atom site in a list of coordinates and atomic displacement factors. Historically, the labels given to atom sites have been chosen to summarize useful information about the atom located at the site. Almost invariably the label contains the symbol of the chemical element or elements occupying the site; there may also be indicators of charge, valence, chemical connectivity, disorder, occupation of a site of crystallographic symmetry or grouping within a component of secondary structure within large molecules. In a CIF, it is formally sufficient that atom-site labels are unique, as all the information about composition, valence, connectivity and so on can be extracted from the data items designed specifically to record this information. However, it is preferable that an atom-site label should summarize the relevant features of the site. Many styles and conventions for labelling atoms are in use in crystallography, so to enable interchange with other crystallographic data file formats, the core dictionary contains a detailed but highly flexible set of rules for constructing and parsing atom-site labels.

Labelling atom sites in crystallography usually serves two distinct purposes: (*a*) to identify the site in the molecule and crystal, and (*b*) to identify the chemical element that occupies that site. The core dictionary makes this distinction clear by defining ATOM_SITE and ATOM_TYPE as separate data categories. The connection between the two categories is made through the equivalence of the data items **_atom_site_type_symbol** (in the ATOM_SITE list) and **_atom_type_symbol** (in the ATOM_TYPE list). Often, however, crystallographers use a single label **_atom_site_label** to define both the site and the chemical species occupying it.

The **_atom_site_label** may be composed of as many as eight separate components; the recommended convention for construction of the string is as follows.

Component 0 [optionally identical to a value of **_atom_type_symbol**] (*mandatory*): A character string containing any character except a blank or an underline, with the proviso that each digit '0'–'9' is used only to designate an oxidation state and, as such, must be followed by a plus '+' or a minus '-' character. It is recommended that the element symbols be used when applicable. Examples of permissible codes are: `Cu`, `Cu2+`, `dummy`, `Fe3+Ni2+`, `S-`, `H*`, `H(SDS)`.

Component 1 [atom number code] (*optional*): This string may contain any alphanumeric character except a blank or an underline, but the first character *must* be a digit '0'–'9' and the second character may not be a plus '+' or a minus '-'. Component 1 is intended primarily to differentiate sites containing the same atom type, but it can be used for any purpose. Examples of combined component 0 and 1 codes are: **C**`1`, **C**`103g28`, **Fe3+**`17b`, **H***`251`, **boron**`2a`, **Ni2+**`2`, **Fe2+Ni2+**`2`, where component 0 is in bold to indicate how these labels are parsed.

Component 2 [residue code] (*optional*): This string may contain any character except a blank or underline. It is intended primarily to give specific structural information such as the molecular fragment or amino-acid type, *e.g.* `C1_gly`, `O1_SO4`. If component 2 is present, it is separated from the concatenated components 0 and 1 with an underline character.

Components 3–7 [sequence, remoteness, chain order, alternate, footnote codes] (*optional*): These strings may contain any character except a blank or an underline. The underline character is used to separate the individual components. The names associated with the separate components suggest their roles in constructing composite labels that match the conventions of site labelling in the PDB format for macromolecular structure files. However, they are not restricted to these functions and may be used in other ways.

Component 0 is normally identical to an **_atom_type_symbol** code in the ATOM_TYPE list. However, if it is not, an **_atom_site_type_symbol** code must appear in the ATOM_TYPE list in order to identify the atom type. In these cases, component 0 may contain any code consistent with the rules given in the dictionary. Thus, component 0 could be `Ca` to identify an alpha carbon, provided that the **_atom_site_type_symbol** is encoded as `C` to indicate that the atom type is carbon.

Multiple occupation of a single atom site by different atom species (compositional disorder) may be handled simply by having multiple values of **_atom_site_label** referring to the same site in the crystal structure. Alternatively, multiple occupancy of an atom site may be denoted by a unique character or characters in component 0 of the atom label, with the ATOM_TYPE list containing the equivalent pseudo element label entry with values that are weighted averages of those for the constituent elements. The proportions of the atom types should then be defined using **_atom_type_description**.

This **_atom_site_label** construction is flexible, visually decipherable and well suited to computer applications. The components can be easily identified and stripped with a single pass, from left to right, along the label string. Note that the underline separators are only used if higher-order components exist. If intermediate components are not used they may be omitted provided the underline separators are retained. For example, the label `C233__ggg` is acceptable and contains the components 0: `C`, 1: `233`, 2: *null* and 3: `ggg`. There is no requirement that the same number of components should be used in each label.

The **_atom_site_label** may be replaced by separate data items specifying the individual components of an atom label; this may be useful for large lists of site coordinates, for example in a macromolecular structure, where site-labelling components follow a systematic convention and where subsets of the atom sites need to be searched for or extracted using individual label components. Such uses are not common in files built with core CIF data names; the mmCIF dictionary identifies substructural components in biological macromolecules by alternative techniques (Section 3.6.7).

There is no comparable fragmentation of the components of **_atom_site_aniso_label**. Where separate lists of anisotropic displacement parameters use complex atom-site labels, either the coordinate list should use **_atom_site_label** alone or the processing software needs to be able to construct a value for **_atom_site_label** from the separate components **_atom_site_label_component_*** in order to test the equivalence between the

Example 3.2.4.3. *Chemical connectivity table; atoms are linked back to atom-site positions through matching values of* **_atom_site_chemical_conn_number** *and* **_chemical_conn_atom_number**.

```
loop_
  _atom_site_label
  _atom_site_chemical_conn_number
  _atom_site_fract_x
  _atom_site_fract_y
  _atom_site_fract_z
  _atom_site_U_iso_or_equiv
  S1 1 0.74799(9)  -0.12482(11) 0.27574(9) 0.0742(3)
  S2 2 1.08535(10)  0.16131(9)  0.34061(9) 0.0741(3)
  N1 3 1.0650(2)   -0.1390(2)   0.2918(2)  0.0500(5)
  C1 4 0.9619(3)   -0.0522(3)   0.3009(2)  0.0509(6)
     # - - - - data truncated for brevity - - - -

loop_
  _chemical_conn_atom_number
  _chemical_conn_atom_type_symbol
  _chemical_conn_atom_display_x
  _chemical_conn_atom_display_y
  _chemical_conn_atom_NCA
  _chemical_conn_atom_NH
1   S    .39  .81   1   0   2   S    .39  .96   2   0
3   N    .14  .88   3   0   4   C    .33  .88   3   0
5   C    .11  .96   2   2   6   C    .03  .96   2   2
     # - - - - data truncated for brevity - - - -
```

Example 3.2.4.4. *Handling of occupational disorder of atom sites.*

```
loop_
  _atom_site_label
  _atom_site_fract_x
  _atom_site_fract_y
  _atom_site_fract_z
  _atom_site_occupancy
  _atom_site_disorder_assembly
  _atom_site_disorder_group
    B2    0.9639(7)  0.6536(5)  0.4464(5)  1     .  .
    F21   0.9411(18) 0.7083(11) 0.5388(9)  0.50  A  1
    F22   1.008(2)   0.5331(6)  0.4747(10) 0.50  A  1
    F23   0.8364(14) 0.6845(17) 0.3951(15) 0.50  A  1
    F24   1.0718(17) 0.6896(14) 0.3764(12) 0.50  A  1
    F21A  0.9727(18) 0.7141(15) 0.5293(13) 0.50  A  2
    F22A  1.0540(15) 0.5401(8)  0.4635(15) 0.50  A  2
    F23A  0.8216(9)  0.6479(15) 0.4461(14) 0.50  A  2
    F24A  1.007(2)   0.7096(17) 0.3476(11) 0.50  A  2
```

labels in the coordinates and anisotropic displacement parameters lists.

While either atom-labelling technique is permitted, it is recommended that the individual label components are *not* used unless there is an overwhelming argument to do so.

Information about the molecular model is sometimes embedded in a labelling convention. In CIF, this information is usually expressed through other data items.

The connectivity of a molecule is described by the CHEMICAL group of categories, and more specifically through the CHEMICAL_CONN_ATOM and CHEMICAL_CONN_BOND categories.

The link between atom sites in the coordinate list and the corresponding atoms in the molecular model is established using the data item **_chemical_conn_atom_number** in the CHEM_CONN_ATOM category, and the data items **_chemical_conn_bond_atom_1** and **_chemical_conn_bond_atom_2** in the CHEMICAL_CONN_BOND category. The values of these data items must match values for the data item **_atom_site_chemical_conn_number** in the ATOM_SITE list. Example 3.2.4.3 shows an extract from a connectivity table; a more complete version of this table is given in the relevant category descriptions in the dictionary.

Note that there is no guarantee that the refined atom-site coordinates that characterize the asymmetric unit will correspond to locations within a single connected molecular species. Crystal symmetry transformations may need to be applied to individual sites in order to map the contents of a connected molecular residue to real space in the unit cell. There is no provision in the CHEMICAL_CONN categories for the specification of these symmetry transformations; thus these higher-order molecular geometries are best described using data items in the GEOM categories, which do allow for the specification of symmetry transformations.

It may also be the case that not all atom positions have been located; this is particularly true for hydrogen atoms, and the data item **_atom_site_attached_hydrogens** is provided for bookkeeping purposes to indicate hydrogen atoms known to be bonded to an atom but whose positions have not been refined (or calculated).

Example 3.2.4.4 shows how the disorder of a group of bonded atoms over a set of atom sites (occupational disorder) is described. In this example of a disordered tetrafluoroborate anion, the data item **_atom_site_disorder_assembly** takes the value A, and the data item **_atom_site_disorder_group** takes the values 1 and 2, indicating the two alternative positions of the disordered group.

The remaining items in this category are clearly described in their individual dictionary entries. However, the now-deprecated data item **_atom_site_refinement_flags** should be mentioned. This was allowed to take values obtained by concatenating one or more of the single-letter flags:

- . no refinement constraints;
- S special-position constraint on site;
- G rigid-group refinement of site;
- R riding-atom site attached to non-riding atom;
- D distance or angle restraint on site;
- T thermal displacement constraints;
- U $U_{\rm iso}$ or U^{ij} restraint (rigid bond);
- P partial occupancy constraint.

These individual flags are listed in the dictionary using the DDL field **_enumeration**, which denotes a list of mutually exclusive permitted values. As concatenation of values is allowed here, dictionary-based software must be modified to handle this data item as a special case. To avoid the need for this in future, the data item was marked as deprecated from version 2.3 of the dictionary, and is replaced by the three separate items **_atom_site_refinement_flags_posn**, ***_adp** and ***_occupancy**. For each of these, the relevant combinations of refinement flags are fully enumerated (for example **_atom_site_refinement_flags_adp** may take any one of the values T, U or TU). This logically separates the different types of refinement constraints or restraints that an author might want to record and allows software to parse the data item.

3.2.4.1.2. *Collections of atom sites*

The data items in this category are as follows:

ATOM_SITES

```
_atom_sites_Cartn_tran_matrix_11
_atom_sites_Cartn_tran_matrix_12
_atom_sites_Cartn_tran_matrix_13
_atom_sites_Cartn_tran_matrix_21
_atom_sites_Cartn_tran_matrix_22
_atom_sites_Cartn_tran_matrix_23
_atom_sites_Cartn_tran_matrix_31
_atom_sites_Cartn_tran_matrix_32
_atom_sites_Cartn_tran_matrix_33
_atom_sites_Cartn_tran_vector_1
_atom_sites_Cartn_tran_vector_2
_atom_sites_Cartn_tran_vector_3
_atom_sites_Cartn_transform_axes
_atom_sites_fract_tran_matrix_11
_atom_sites_fract_tran_matrix_12
_atom_sites_fract_tran_matrix_13
```

```
_atom_sites_fract_tran_matrix_21
_atom_sites_fract_tran_matrix_22
_atom_sites_fract_tran_matrix_23
_atom_sites_fract_tran_matrix_31
_atom_sites_fract_tran_matrix_32
_atom_sites_fract_tran_matrix_33
_atom_sites_fract_tran_vector_1
_atom_sites_fract_tran_vector_2
_atom_sites_fract_tran_vector_3
_atom_sites_solution_hydrogens
_atom_sites_solution_primary
_atom_sites_solution_secondary
_atom_sites_special_details
```

This category records information that applies collectively to the atom sites of the structural model. At present, the topics covered are the transformation matrix between Cartesian and cell fractional coordinates, and the methods used to locate the initial atom sites. **_atom_sites_solution_primary** describes how the first atom sites were determined, **_atom_sites_solution_secondary** describes how the remaining non-hydrogen sites were located and **_atom_sites_solution_hydrogens** describes how hydrogen atoms were located. The codes that are allowed for each of these refer to distinct solution methods, and at present only the seven formal values listed below are provided (although other values might be added in the future):

`difmap`	difference-electron-density map;
`vecmap`	real-space vector search;
`heavy`	heavy-atom method;
`direct`	structure-invariant direct methods;
`geom`	inferred from neighbouring sites;
`disper`	anomalous-dispersion techniques;
`isomor`	isomorphous structure methods.

3.2.4.1.3. *Atom types*

The data items in this category are as follows:

ATOM_TYPE

- **_atom_type_symbol**

```
_atom_type_analytical_mass_%
_atom_type_description
_atom_type_number_in_cell
_atom_type_oxidation_number
_atom_type_radius_bond
_atom_type_radius_contact
_atom_type_scat_Cromer_Mann_a1
_atom_type_scat_Cromer_Mann_a2
_atom_type_scat_Cromer_Mann_a3
_atom_type_scat_Cromer_Mann_a4
_atom_type_scat_Cromer_Mann_b1
_atom_type_scat_Cromer_Mann_b2
_atom_type_scat_Cromer_Mann_b3
_atom_type_scat_Cromer_Mann_b4
_atom_type_scat_Cromer_Mann_c
_atom_type_scat_dispersion_imag
_atom_type_scat_dispersion_real
_atom_type_scat_dispersion_source
_atom_type_scat_length_neutron
_atom_type_scat_source
_atom_type_scat_versus_stol_list
```

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

The data items in this category record details about the atomic species associated with each occupied atom site in the structural model. While these will typically be standard properties of the naturally occurring chemical elements, they may also be synthetic atom types, for example in cases where a single atom site may be occupied with partial occupancies by atoms of different elements.

As mentioned in Section 3.2.4.1.1, there are two ways of dealing with such a case: the same location in the coordinate list may be populated by multiple entries, each for an atom of a particular element with an associated occupancy fraction; or a single entry may be made for the synthetic atom type, the properties of which are described fully in the ATOM_TYPE list.

Example 3.2.4.5. *Reference to atomic scattering factors.*

```
loop_
  _atom_type_symbol
  _atom_type_oxidation_number
  _atom_type_number_in_cell
  _atom_type_scat_dispersion_real
  _atom_type_scat_dispersion_imag
  _atom_type_scat_source
    C  0  72  .017  .009
              International_Tables_Vol_IV_Table_2.2B
    H  0  100  0     0
              International_Tables_Vol_IV_Table_2.2B
    O  0  12  .047  .032
              International_Tables_Vol_IV_Table_2.2B
    N  0  4   .029  .018
              International_Tables_Vol_IV_Table_2.2B
```

Each different atom type has a unique **_atom_type_symbol** identifier. In principle, this could be any string of characters, but the dictionary recommends certain conventions to encourage compatibility with the atom-site labelling rules. It is recommended that the identifier be the normal chemical element symbol when the atom type is a pure element. If some other labelling is used, the identifier may be composed of any character except an underline, with the additional proviso that digits designate an oxidation state and must be followed by a '+' or '-' character.

The data item **_atom_type_scat_versus_stol_list** can be used to give a table of scattering factors as a function of $(\sin\theta)/\lambda$. This is a text field with no specified internal structure, except the suggestion that it is well commented and the lists should be regularly formatted. However, it is generally enough to list the atomic scattering factors of each element and to provide a reference to the source of the values, as in Example 3.2.4.5.

3.2.4.2. Chemical identification and connectivity information

The categories describing chemical identity and connectivity are as follows:

CHEMICAL group

Chemical identification (§3.2.4.2.1)

CHEMICAL

CHEMICAL_FORMULA

Chemical connectivity (§3.2.4.2.2)

CHEMICAL_CONN_ATOM

CHEMICAL_CONN_BOND

As indicated in Section 3.2.4.1.1, the chemical interpretation of the coordinate list of regions of significant electron density is not always easy. Occupational and compositional disorder, symmetry-equivalent locations, and unrefined atom sites all contribute to the difficulties, but it is usually possible in modern studies to construct a sensible chemical model. The CHEMICAL category group provides the data names needed to describe the chemical identity and properties of the material characterized in the structural study.

3.2.4.2.1. *Chemical identification*

The data items in these categories are as follows:

(*a*) CHEMICAL

```
_chemical_absolute_configuration
_chemical_compound_source
_chemical_melting_point
_chemical_melting_point_gt
_chemical_melting_point_lt
_chemical_name_common
_chemical_name_mineral
_chemical_name_structure_type
_chemical_name_systematic
```

```
_chemical_optical_rotation
_chemical_properties_biological
_chemical_properties_physical
_chemical_temperature_decomposition
_chemical_temperature_decomposition_gt
_chemical_temperature_decomposition_lt
_chemical_temperature_sublimation
_chemical_temperature_sublimation_gt
_chemical_temperature_sublimation_lt
```

(*b*) CHEMICAL_FORMULA

```
_chemical_formula_analytical
_chemical_formula_iupac
_chemical_formula_moiety
_chemical_formula_structural
_chemical_formula_sum
_chemical_formula_weight
_chemical_formula_weight_meas
```

The CHEMICAL category itself deals with the large-scale chemical properties of the compound from which the crystal under study was formed: its various formal and common names, its source, melting point, decomposition and sublimation temperatures (as experimentally determined values, or as upper or lower possible values if not measured directly), its biological or physical properties, and where applicable the absolute configuration and optical rotation.

The optical rotation in solution may be reported using the data name **_chemical_optical_rotation** by an expression of the form

$$[\alpha]_W^T = \pm\frac{100\alpha}{lc} \qquad (c = CONC, \mathrm{SOLV}),$$

where $[\alpha]_W^T$ is the signed optical rotation in degrees at temperature T and wavelength labelled by code W, l is the length of the optical cell, *CONC* is the concentration of the solution (given as the mass of the substance in g in a standard 100 ml of solution), and SOLV is the chemical formula of the solvent. This can be marked up within the constraints of the ASCII character set to which CIF is restricted as `[\a]^25^~D~ = +108 (c = 3.42, CHCl~3~)`, where the measurement is taken using the *D* line of the atomic spectrum of sodium.

Data items in the CHEMICAL_FORMULA category describe the chemical formula and formula mass of the compound under study. The quoted formula must reflect the overall stoichiometry of the crystal under study, and must, when multiplied by the Z value **_cell_formula_units_Z**, account for the total contents of the unit cell.

A number of data names are provided to account for different conventions in the presentation of chemical formulae. **_chemical_formula_analytical** is appropriate for a gross formula determined by standard chemical analysis, including all trace elements identified in the sample. Standard uncertainties on the proportions of elements present are acceptable, *e.g.*

```
_chemical_formula_analytical 'Fe2.45(2) Ni1.60(3) S4'
```

_chemical_formula_sum is another aggregate formula, in which all discrete bonded residues and ions are summed over the constituent elements. Where appropriate, the formulae of separate residues of a complex may be described by **_chemical_formula_moiety**, in which the formula for each moiety is supplied as a sum of the individual elements within the moiety, or by **_chemical_formula_structural**, in which sub-components within individual moieties are further identified, so that the overall expression permits the identification of particular bonded groups. Within these formula expressions, certain rules must be observed to allow parsing by software. The final data item relating to the chemical formula, **_chemical_formula_iupac**, is for formulae that are constructed according to the rules of the International Union for Pure and Applied Chemistry.

The ordering and notation rules are explained n detail in the dictionary, but are repeated here for convenience. Within each group of atoms for which a formula is present:

(i) only recognized element symbols may be used;

(ii) each element symbol is followed by a 'count' number ('1' is implicit and may be omitted);

(iii) a space or parenthesis must separate each cluster of (element symbol + count);

(iv) where a group of elements is enclosed in parentheses, the multiplier for the group must follow the closing parentheses. That is, all element and group multipliers are assumed to be printed as subscripted numbers. (An exception to this rule exists for **_chemical_formula_moiety**, where pre- and post-multipliers are permitted for molecular units.)

(v) Unless the elements are ordered in a manner that corresponds to their chemical structure, as in **_chemical_formula_structural**, the order of the elements within any group or moiety depends on whether or not carbon is present. If carbon is present, the order should be: C, then H, then the other elements in alphabetical order of their symbol. If carbon is not present, the elements are listed purely in alphabetic order of their symbol. This is the 'Hill' system used by *Chemical Abstracts*. This ordering is used in **_chemical_formula_moiety** and **_chemical_formula_sum**.

For **_chemical_formula_moiety** some additional rules apply:

(i) Moieties are separated by commas, ','.

(ii) The order of elements within a moiety follows the general rules outlined above as the 'Hill' system.

(iii) Parentheses are *not* used within moieties but may surround a moiety. Parentheses may not be nested.

(iv) Charges should be placed at the end of the moiety. The charge '+' or '-' may be preceded by a numerical multiplier and should be separated from the last (element symbol + count) by a space. Pre- or post-multipliers may be used for individual moieties.

Example 3.2.4.6 illustrates the differences between some of these data items.

3.2.4.2.2. *Chemical connectivity*

The data items in these categories are as follows:

(*a*) CHEMICAL_CONN_ATOM

```
• _chemical_conn_atom_number
  _chemical_conn_atom_charge
  _chemical_conn_atom_display_x
  _chemical_conn_atom_display_y
  _chemical_conn_atom_NCA
  _chemical_conn_atom_NH
  _chemical_conn_atom_type_symbol
```

(*b*) CHEMICAL_CONN_BOND

```
• _chemical_conn_bond_atom_1
        → _chemical_conn_atom_number
• _chemical_conn_bond_atom_2
        → _chemical_conn_atom_number
  _chemical_conn_bond_type
```

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 CHEMICAL_CONN_ATOM category labels the chemical atoms in a connected representation of the molecular species and can also give the coordinates for the atoms in a two-dimensional chemical diagram (Example 3.2.4.7). Each atom may also carry an indication of the number of connected non-hydrogen atoms (**_NCA**) and the number of hydrogen atoms (**_NH**) to which it

Example 3.2.4.6. *Different representations of a chemical formula.*

```
_chemical_formula_iupac    '[Mo (C O)4 (C18 H33 P)2]'
_chemical_formula_moiety    'Mo,4(C O),2(C18 H33 P)'
_chemical_formula_structural
                          '((C O)4 (P (C6 H11)3)2)Mo'
_chemical_formula_sum              'C40 H66 Mo O4 P2'
```

Example 3.2.4.7. *Representation of a two-dimensional chemical diagram.*

```
loop_
     _chemical_conn_atom_number
     _chemical_conn_atom_type_symbol
     _chemical_conn_atom_display_x
     _chemical_conn_atom_display_y
     _chemical_conn_atom_NCA
     _chemical_conn_atom_NH
        1   S    .39  .81   1   0
        2   S    .39  .96   2   0
        3   N    .14  .88   3   0
        4   C    .33  .88   3   0
        5   C    .11  .96   2   2
        6   C    .03  .96   2   2
        7   C    .03  .80   2   2
        8   C    .11  .80   2   2
        9   S    .54  .81   1   0
        10  S    .54  .96   2   0
        11  N    .80  .88   3   0
        12  C    .60  .88   3   0
        13  C    .84  .96   2   2
        14  C    .91  .96   2   2
        15  C    .91  .80   2   2
        16  C    .84  .80   2   2
```

is connected. Together with the CHEMICAL_CONN_BOND category, the data items in the CHEMICAL_CONN_ATOM category provide a basic description of the chemical structure. Although the description of the chemical structure provided in these two categories is not as extensive as the information that may be conveyed in a molecular information file (Chapter 2.4), it should allow a substructure to be searched for in a suitable database.

The CHEMICAL_CONN_BOND category lists pairs of atoms that contribute to chemical bonds and describes the nature of the bond between them (Example 3.2.4.8). Taken with data items in the CHEMICAL_CONN_ATOM category, data items in this category complete the basic description of a molecular entity.

Bond types are assigned from a list that specifies single, double, triple, quadruple, aromatic, polymeric, delocalized double and π bonds. These are not intended to cover all possible cases, but to characterize a molecular model suitable for database substructure searching.

3.2.4.3. Molecular or packing geometry

The categories describing geometry are as follows:

GEOM group

- GEOM
- GEOM_ANGLE
- GEOM_BOND
- GEOM_CONTACT
- GEOM_HBOND
- GEOM_TORSION

The molecular and packing geometry can be calculated fully given the unit-cell parameters, the space group and a list of atom sites. Therefore, all the information about geometry in the GEOM category group is derivative. However, it is useful to record it within the file both as a check on the primary information stored in other categories and as a method for flagging values to be published.

Example 3.2.4.8. *Bond types in a chemical connectivity table.*

```
loop_
     _chemical_conn_bond_atom_1
     _chemical_conn_bond_atom_2
     _chemical_conn_bond_type
       4     1     doub     4     3     sing
       4     2     sing     5     3     sing
       6     5     sing     7     6     sing
       8     7     sing     8     3     sing
       10    2     sing     12    9     doub
       12    11    sing     12    10    sing
       13    11    sing     14    13    sing
       15    14    sing     16    15    sing
       16    11    sing     17    5     sing
       18    5     sing     19    6     sing
       20    6     sing     21    7     sing
       22    7     sing     23    8     sing
       24    8     sing     25    13    sing
       26    13    sing     27    14    sing
       28    14    sing     29    15    sing
       30    15    sing     31    16    sing
       32    16    sing
```

3.2.4.3.1. *Contents of the geometry-related categories*

The data items in these categories are as follows:

(*a*) GEOM

```
_geom_special_details
```

(*b*) GEOM_ANGLE

- _geom_angle_atom_site_label_1 → _atom_site_label
- _geom_angle_atom_site_label_2 → _atom_site_label
- _geom_angle_atom_site_label_3 → _atom_site_label
- _geom_angle_site_symmetry_1
- _geom_angle_site_symmetry_2
- _geom_angle_site_symmetry_3

```
_geom_angle
_geom_angle_publ_flag
```

(*c*) GEOM_BOND

- _geom_bond_atom_site_label_1 → _atom_site_label
- _geom_bond_atom_site_label_2 → _atom_site_label
- _geom_bond_site_symmetry_1
- _geom_bond_site_symmetry_2

```
_geom_bond_distance
_geom_bond_publ_flag
_geom_bond_valence
```

(*d*) GEOM_CONTACT

- _geom_contact_atom_site_label_1 → _atom_site_label
- _geom_contact_atom_site_label_2 → _atom_site_label
- _geom_contact_site_symmetry_1
- _geom_contact_site_symmetry_2

```
_geom_contact_distance
_geom_contact_publ_flag
```

(*e*) GEOM_HBOND

- _geom_hbond_atom_site_label_A → _atom_site_label
- _geom_hbond_atom_site_label_D → _atom_site_label
- _geom_hbond_atom_site_label_H → _atom_site_label
- _geom_hbond_site_symmetry_A
- _geom_hbond_site_symmetry_D
- _geom_hbond_site_symmetry_H

```
_geom_hbond_angle_DHA
_geom_hbond_distance_DA
_geom_hbond_distance_DH
_geom_hbond_distance_HA
_geom_hbond_publ_flag
```

(*f*) GEOM_TORSION

- `_geom_torsion_atom_site_label_1`
  → `_atom_site_label`
- `_geom_torsion_atom_site_label_2`
  → `_atom_site_label`
- `_geom_torsion_atom_site_label_3`
  → `_atom_site_label`
- `_geom_torsion_atom_site_label_4`
  → `_atom_site_label`
- `_geom_torsion_site_symmetry_1`
- `_geom_torsion_site_symmetry_2`
- `_geom_torsion_site_symmetry_3`
- `_geom_torsion_site_symmetry_4`

  `_geom_torsion`
  `_geom_torsion_publ_flag`

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. ***_symmetry_**** *items have a default value and may be omitted from the list. The arrow (→) is a reference to a parent data item.*

Most categories within this group record distances or angles specified by atom-site labels and are well characterized. The GEOM category currently provides the single data name **_geom_special_details** in which any other details of the geometry that an author considers noteworthy may be stored. Examples of information that might be stored in this data item are least-squares equations of planes, out-of-plane distances, dihedral angles between planes and general comments about the calculation of standard uncertainties.

A subtlety in the geometry-related categories arises from the need to record geometric relationships that involve atoms that are not listed in the ATOM_SITE coordinate list, but that can be derived from the coordinates in this list by the application of a crystallographic symmetry transformation. Thus atom sites in the geometry lists are identified both by their atom-site labels (which must identically match one of the entries in the ATOM_SITE list) *and* by the code for the symmetry transformation that has been applied to the initial location. Since the atom-site labels may refer to atoms in their original location as well as to atoms in symmetry-related locations, the formal key for these categories involves the site labels as well as the symmetry codes. However, in many cases (as discussed further below) the symmetry codes may be absent from a list, and a parser must supply suitable default or null values for the missing components when constructing or checking a complete key.

In many cases, interest is focused on intramolecular distances and angles, and on intramolecular contacts within a single asymmetric unit. In such cases, the geometry lists would contain only atoms listed explicitly in the ATOM_SITE list and the symmetry codes all refer trivially to the identity transformation.

The examples in this section demonstrate various ways of handling geometry lists with trivial or non-trivial symmetry transformations. In Example 3.2.4.9, showing treatment of bond angles, the relevant data items (**_geom_angle_site_symmetry_***) are absent, which is one method for indicating the identity transformation. Dictionary validation software must therefore be able to handle both the presence and absence of these components of the formal category key.

The symmetry transformations in this and related categories take the form of codes *'n klm'* or *n_klm*, where *n* refers to the symmetry operation that is applied to the coordinates stored in **_atom_site_fract_x**, **_atom_site_fract_y** and **_atom_site_fract_z**. The value of *n* must match a number given in **_symmetry_equiv_pos_site_id**. *k*, *l* and *m* refer to the translations that are subsequently applied to the symmetry-transformed coordinates to generate the atom used in calculating the contact. These translations (x, y, z) are related to (k, l, m) by

Example 3.2.4.9. *List of bond angles.*

```
loop_
     _geom_angle_atom_site_label_1
     _geom_angle_atom_site_label_2
     _geom_angle_atom_site_label_3
     _geom_angle
     _geom_angle_publ_flag
     C2  O1  C5   111.6(2)  yes
     O1  C2  C3   110.9(2)  yes
     O1  C2  O21  122.2(3)  yes
     C3  C2  O21  127.0(3)  yes
     C2  C3  N4   101.3(2)  yes
     C2  C3  C31  111.3(2)  yes
     C2  C3  H3   107(1)    no
     N4  C3  C31  116.7(2)  yes
```

Example 3.2.4.10. *List of bonds.*

```
loop_
     _geom_bond_atom_site_label_1
     _geom_bond_atom_site_label_2
     _geom_bond_distance
     _geom_bond_site_symmetry_1
     _geom_bond_site_symmetry_2
     _geom_bond_publ_flag
      O1  C2   1.342(4)  1_555  1_555  yes
      O1  C5   1.439(3)  1_555  1_555  yes
      C2  C3   1.512(4)  1_555  1_555  yes
      C2  O21  1.199(4)  1_555  1_555  yes
      C3  N4   1.465(3)  1_555  1_555  yes
      C3  C31  1.537(4)  1_555  1_555  yes
      C3  H3   1.00(3)   1_555  1_555  no
      N4  C5   1.472(3)  1_555  1_555  yes
```

$$k = 5 + x, \qquad l = 5 + y, \qquad m = 5 + z.$$

By adding 5 to the translations, the use of negative numbers is avoided. As an example, the symmetry code 7_645 means that the symmetry operation with label '7' in the **_symmetry_equiv_pos_site_id** list is applied and the resulting position is translated $+1.0 \times a$ along the x axis, $-1.0 \times b$ along the y axis and $0.0 \times c$ along the z axis, where a, b and c are the unit-cell edges.

List entries with a **_geom_angle_publ_flag** value of `yes` are those that should be published.

The GEOM_BOND category records intramolecular bond distances. In Example 3.2.4.10, all the atoms are untransformed and are at the positions given in the ATOM_SITE list. The symmetry code is 1_555, where the trivial symmetry operation x, y, z is numbered '1' by **_symmetry_equiv_pos_site_id**.

The GEOM_CONTACT category records nonbonded interatomic contacts. In Example 3.2.4.11, all the atoms are untransformed and are at the positions given in the ATOM_SITE list, and therefore the symmetry codes all have the value '.' (meaning 'inapplicable'). This is another method for indicating the identity transformation.

The GEOM_HBOND category records details about hydrogen bonds. Unlike other categories in the GEOM group, the GEOM_HBOND category records information about both distances and angles, including donor–acceptor, donor–hydrogen and acceptor–hydrogen distances and the included angle at the hydrogen-atom site (see Example 3.2.4.12). The comments above about the interpretation of symmetry codes and their relevance in the formal assignment of the category key also apply to this category.

Note that, strictly speaking, this category should only be populated if coordinates for the hydrogen atom are available (because the mandatory component of the category key **_geom_hbond_**

Example 3.2.4.11. *List of nonbonded interatomic contacts.*

```
loop_
     _geom_contact_atom_site_label_1
     _geom_contact_atom_site_label_2
     _geom_contact_distance
     _geom_contact_site_symmetry_1
     _geom_contact_site_symmetry_2
     _geom_contact_publ_flag
     O(1)  O(2)     2.735(3)  .  .  yes
     H(O1) O(2)     1.82      .  .  no
```

Example 3.2.4.12. *List of hydrogen-bond distances and angles.*

```
loop_
     _geom_hbond_atom_site_label_D
     _geom_hbond_atom_site_label_H
     _geom_hbond_atom_site_label_A
     _geom_hbond_distance_DH
     _geom_hbond_distance_HA
     _geom_hbond_distance_DA
     _geom_hbond_angle_DHA
     _geom_hbond_publ_flag
N6  HN6 OW  0.888(8) 1.921(12) 2.801(8) 169.6(8) yes
OW  HO2 O7  0.917(6) 1.923(12) 2.793(8) 153.5(8) yes
OW  HO1 N10 0.894(8) 1.886(11) 2.842(8) 179.7(9) yes
```

Example 3.2.4.13. *List of torsion angles.*

```
    loop_
     _geom_torsion_atom_site_label_1
     _geom_torsion_atom_site_label_2
     _geom_torsion_atom_site_label_3
     _geom_torsion_atom_site_label_4
     _geom_torsion
     _geom_torsion_site_symmetry_1
     _geom_torsion_site_symmetry_2
     _geom_torsion_site_symmetry_3
     _geom_torsion_site_symmetry_4
     _geom_torsion_publ_flag
C(9)  O(2) C(7)  C(2)    71.8(2)  .  .  .  .      yes
C(7)  O(2) C(9)  C(10) -168.0(3)  .  .  .  2_666  yes
C(10) O(3) C(8)  C(6)  -167.7(3)  .  .  .  .      yes
C(8)  O(3) C(10) C(9)   -69.7(2)  .  .  .  2_666  yes
O(1)  C(1) C(2)  C(3)  -179.5(4)  .  .  .  .      no
O(1)  C(1) C(2)  C(7)    -0.6(1)  .  .  .  .      no
```

`atom_site_label_H` needs a parent label in the atom-site list). In practice, hydrogen bonds can be assumed between donor atoms and acceptors even if the hydrogen atom is not specifically located.

The items in the GEOM_TORSION category describe the torsion angle in degrees generated for the bonded sequence of four atom sites identified by the `_geom_torsion_atom_site_label_*` codes. As with other geometry-specific site labels, these must match labels specified as `_atom_site_label` in the atom list. The torsion angle definition is that of Klyne & Prelog (1960).

Example 3.2.4.13 includes two sites that have been generated by crystallographic symmetry operations and lattice translations from the parent sites in the atom list.

3.2.4.4. Symmetry and space-group information

The categories describing symmetry are as follows:

SYMMETRY group

Original symmetry categories (§3.2.4.4.1)

SYMMETRY

SYMMETRY_EQUIV

Replacement symmetry categories (§3.2.4.4.2)

SPACE_GROUP

SPACE_GROUP_SYMOP

The SPACE_GROUP and older SYMMETRY categories contain information about the symmetry of the crystal; specifically the space group and the symmetry-equivalent positions for that space group. More information about the symmetry is available in the symCIF dictionary described in Chapter 3.8 and presented in Chapter 4.7. The categories SPACE_GROUP and SPACE_GROUP_SYMOP were imported from symCIF in version 2.3 of the core dictionary, and are intended to replace the SYMMETRY and SYMMETRY_EQUIV categories. In most cases, there are strict equivalences between data items in the two sets. The new categories have been adopted for greater compatibility with future expansions to the symmetry CIF dictionary, and to correct some potentially misleading practices in the original categories. Although all the data items in SYMMETRY and SYMMETRY_EQUIV_POS are now formally marked as deprecated, it is likely that the older data items will remain in circulation for some time.

3.2.4.4.1. *Data items in SYMMETRY and related categories*

The data items in these categories are as follows:

(*a*) SYMMETRY

† `_symmetry_cell_setting`
† `_symmetry_Int_Tables_number`
† `_symmetry_space_group_name_H-M`
† `_symmetry_space_group_name_Hall`

(*b*) SYMMETRY_EQUIV

†• `_symmetry_equiv_pos_site_id`
† `_symmetry_equiv_pos_as_xyz`

The bullet (•) *indicates a category key. In practice* `_symmetry_equiv_pos_site_id` *is often absent from older CIFs. The dagger* (†) *indicates a deprecated item, which should not be used in the creation of new CIFs.*

The data items in the SYMMETRY category (now superseded by SPACE_GROUP) were used to record the space group. The Hermann–Mauguin (H-M) symbol was given by `_symmetry_space_group_name_H-M`. The dictionary definition recommended the use of the 'full' H-M symbol as listed in *International Tables for Crystallography* Volume A, but was not explicit about the meaning of 'full'. The dictionary examples showed short-form symbols expanded to a complete representation of individual symmetry elements; thus *Pnnn* would be given as `'P 2/n 2/n 2/n'`, and the monoclinic space group $P2_1/m$ would be given as `'P 1 21/m 1'` for the *b*-axis unique setting or `'P 1 1 21/m'` for the *c*-axis unique setting.

In practice, abbreviated symbols were often used, following conventions established over many years; thus `'P 21/m'` was often given as the Hermann–Mauguin symbol when the 'usual' *b* setting of a monoclinic cell had been chosen. It is recommended that these conventions should continue to be followed when the new data item `_space_group_name_H-M_alt` is used instead.

The dictionary examples also suggested concise ways of indicating the origin choice within the `_symmetry_space_group_name_H-M` field; since there is no formal description of how to do this, different authors used different wording. Hence, `_symmetry_space_group_name_H-M` was always best considered as a container for the representation of the space group that would appear in a published article, and not as a machine-readable source of information about the crystallographic symmetry.

The two mechanisms for conveying the symmetry transformations in a fully machine-readable form were the Hall symbol `_symmetry_space_group_name_Hall` (Hall, 1981; Hall & Grosse-Kunstleve, 2001) and a complete listing of the symmetry operations using data items in the SYMMETRY_EQUIV category.

The data item `_symmetry_cell_setting` indicates the crystal system, not (as suggested by its name) the setting used.

Example 3.2.4.14. *A list of symmetry-equivalent positions.*

```
loop_
     _symmetry_equiv_pos_site_id
     _symmetry_equiv_pos_as_xyz
        1     x,y,z
        2     1/2-x,-y,1/2+z
        3     1/2+x,1/2-y,-z
        4     -x,1/2+y,1/2-z
```

The SYMMETRY_EQUIV category, now superseded by SPACE_GROUP_SYMOP, provided a list of symmetry-equivalent positions in algebraic notation. Formally, **_symmetry_equiv_pos_site_id** acted as a category key, with any arbitrary numeric value that uniquely identifies each operator. Historically, the earliest versions of the core dictionary did not have such an identifier at all and the separate equivalent positions were indexed by their position in the **_symmetry_equiv_pos_as_xyz** list. This interpretation was vulnerable to inadvertent re-ordering of the list of equivalent positions, and for this reason, as well as to satisfy the formal need for a category key, **_symmetry_equiv_pos_site_id** was added (Example 3.2.4.14). For compatibility with software that was written to handle the earlier arrangement, it is recommended that **_symmetry_equiv_pos_site_id** gives sequential integer labels, starting with 1, to the equivalent positions in the sequence in which they appear in the CIF.

Note that the **_symmetry_equiv_pos_as_xyz** list must contain *all* symmetry-equivalent positions of the space group, including those generated by lattice centring and a centre of symmetry, if present.

3.2.4.4.2. *Data items in SPACE_GROUP and related categories*

Data items in these categories are as follows:

(*a*) SPACE_GROUP

_space_group_crystal_system
_space_group_id
_space_group_IT_number
_space_group_name_Hall
_space_group_name_H-M_alt

(*b*) SPACE_GROUP_SYMOP

• **_space_group_symop_id**
_space_group_symop_operation_xyz
_space_group_symop_sg_id

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

The data items in the SPACE_GROUP category record the space group and crystal system. They recognize the common practice of supplying the space group in Hermann–Mauguin notation, though the H-M symbol does not contain complete information about the symmetry and the space-group origin. **_space_group_name_H-M_alt** allows the use of any legitimate H-M symbol as listed in *International Tables for Crystallography* Volume A or derived by similar principles. It does not give rigorous direction on how the symbols should be presented. It is recommended that the use of this symbol in CIFs containing articles for publication should follow the guidelines for **_symmetry_space_group_name_H-M** (Section 3.2.4.4.1).

Because a given space-group type may be described by more than one Hermann–Mauguin symbol, the space-group type should be specified by the use of **_space_group_IT_number**.

Two mechanisms exist for conveying fully machine-readable descriptions of the symmetry transformations relevant to the space group and setting. The first is the Hall symbol (Hall, 1981; Hall & Grosse-Kunstleve, 2001), which uniquely defines the space group and its reference to a particular coordinate system; it is specified in the data item **_space_group_name_Hall**. Alternatively, the symmetry operations may be listed in full using data items in the SYMMETRY_EQUIV category.

Example 3.2.4.15. *A list of symmetry operators using data items from the SPACE_GROUP_SYMOP category.*

```
loop_
     _space_group_symop_id
     _space_group_symop_operation_xyz
        1     x,y,z
        2     1/2-x,-y,1/2+z
        3     1/2+x,1/2-y,-z
        4     -x,1/2+y,1/2-z
```

The SPACE_GROUP_SYMOP category provides a list of the symmetry operators for a space group in algebraic notation. It replaces the category SYMMETRY_EQUIV_POS. Unlike the older category, where in practice the category key could be omitted from listings (and must therefore be generated implicitly by parsing software), the category key **_space_group_symop_id** *must* be given. See Example 3.2.4.15, which may be compared with Example 3.2.4.14.

3.2.4.5. Bond-valence information

Categories describing bond valences are as follows:

VALENCE group
  VALENCE_PARAM
  VALENCE_REF

Data items in these categories are as follows:

(*a*) VALENCE_PARAM

_valence_param_atom_1
_valence_param_atom_1_valence
_valence_param_atom_2
_valence_param_atom_2_valence
_valence_param_B
_valence_param_details
_valence_param_id
_valence_param_ref_id
  → **_valence_ref_id**
_valence_param_Ro

(*b*) VALENCE_REF

_valence_ref_id
_valence_ref_reference

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

The data items in this category group relate to bond valences, which are widely used in inorganic crystallography to confirm and analyse the results of crystal structure determinations. Bond valences are determined from the bond lengths and have the useful property that their sum around any atom is equal to the atom valence (formal charge). They are increasingly being published with bond lengths. The data item **_geom_bond_valence** in the GEOM_BOND category allows the bond valence to be associated with the bond length.

The two categories discussed here list the parameters used to calculate the bond valences and their literature sources. These items might also be published, particularly where there is some uncertainty about the appropriate parameters to use.

The data items in the VALENCE_PARAM category define the parameters used for calculating bond valences from bond lengths. In addition to the parameters, a pointer to the reference for the source of the parameters (in VALENCE_REF) is given (Example 3.2.4.16).

Example 3.2.4.16. *A list of bond-valence parameters.*

```
loop_
     _valence_param_atom_1
     _valence_param_atom_1_valence
     _valence_param_atom_2
     _valence_param_atom_2_valence
     _valence_param_Ro
     _valence_param_B
     _valence_param_ref_id
     _valence_param_details
      Cu 2 O -2 1.679 0.37 a .
      Cu 2 O -2 1.649 0.37 j .
      Cu 2 N -3 1.64  0.37 m '2-coordinate N'
      Cu 2 N -3 1.76  0.37 m '3-coordinate N'
loop_
     _valence_ref_id
     _valence_ref_reference
a
'Brown & Altermatt (1985), Acta Cryst. B41, 244-247'
j
'Liu & Thorp (1993), Inorg. Chem. 32, 4102-4205'
m
; See, Krause & Strub (1998), Inorg. Chem.
  37, 5369-5375'
;
```

3.2.5. Publication

As an archival file format, CIF is well suited to the complete documentation of a structural study and the categories described in this section provide data items suitable for the generation of a fully documented report, either as an informal laboratory notebook document or as a formal published article.

3.2.5.1. Literature citations

The categories describing literature citations are as follows:

CITATION group
- CITATION
- CITATION_AUTHOR
- CITATION_EDITOR

The entries in the CITATION category group provide a set of data items suitable for the structured recording of references to the literature. At present, they are designed for the storage and retrieval of information about journal articles and individual chapters in books. They do not currently cover conference proceedings, pamphlets, preprints, theses or other kinds of publication. Reference lists are usually requested by journals that accept articles in CIF format as a single text field in `_publ_section_references`, but the categories in the CITATION group may become more useful for storing citation lists in the future, especially if converters become available to and from other bibliographic formats such as EndNote and BibTEX.

Data items in these categories are as follows:

(*a*) CITATION
- `_citation_id`
  `_citation_abstract`
  `_citation_abstract_id_CAS`
  `_citation_book_id_ISBN`
  `_citation_book_publisher`
  `_citation_book_publisher_city`
  `_citation_book_title`
  `_citation_coordinate_linkage`
  `_citation_country`
  `_citation_database_id_CSD`
  `_citation_database_id_Medline`
  `_citation_journal_abbrev`
  `_citation_journal_full`
  `_citation_journal_id_ASTM`
  `_citation_journal_id_CSD`
  `_citation_journal_id_ISSN`
  `_citation_journal_issue`
  `_citation_journal_volume`
  `_citation_language`
  `_citation_page_first`
  `_citation_page_last`
  `_citation_special_details`
  `_citation_title`
  `_citation_year`

(*b*) CITATION_AUTHOR
- `_citation_author_citation_id`
  → `_citation_id`
  `_citation_author_name`
  `_citation_author_ordinal`

(*c*) CITATION_EDITOR
- `_citation_editor_citation_id`
  → `_citation_id`
  `_citation_editor_name`
  `_citation_editor_ordinal`

The bullet (•) indicates a category key. The arrow (→) is a reference to a parent data item.

The CITATION category provides the bulk of the information about individual citations. `_citation_id` provides a link to the CITATION_AUTHOR and CITATION_EDITOR categories, where multiple authors, and, if appropriate, multiple editors are listed.

Example 3.2.5.1 shows how a fully populated citation list is structured across these categories.

The authors of a cited reference are listed using items from the CITATION_AUTHOR category. The value of `_citation_author_citation_id` must match a value of `_citation_id` in the CITATION category, and this data item forms the link between the authors and the citations. `_citation_author_ordinal` is used to record the order in which the authors are listed.

The editors of a cited reference are listed using items from the CITATION_EDITOR category. The value of `_citation_editor_citation_id` must match a value of `_citation_id` in the CITATION category, and this data item forms the link between the editors and the citations. `_citation_editor_ordinal` is used to record the order in which the editors are listed.

Example 3.2.5.1. *A structured bibliographic reference list.*

```
loop_
     _citation_id
     _citation_title
     _citation_page_first
     _citation_page_last
     _citation_year
     _citation_journal_abbrev
     _citation_journal_volume
     _citation_journal_id_ISSN
      1
;     Angle calculations for 3- and 4-circle X-ray
      and neutron diffractometers
;
      457  464  1967  'Acta Cryst.'  22  0365-110X
      2
     'Space-group notation with an explicit origin'
      517  525  1981  'Acta Cryst. Section A'  37
      0108-7673
      3 ? 521  523  1960  'Experientia' 16 ?

loop_
     _citation_author_citation_id
     _citation_author_name
      1   'Busing, W. R.'
      1   'Levy, H. A.'
      2   'Hall, S. R.'
      3   'Klyne, W.'
      3   'Prelog, V.'
```

3.2.5.2. Citation of software packages

The single category describing software citations is as follows:

COMPUTING group
  COMPUTING

Data items in this category are as follows:

COMPUTING
  _computing_cell_refinement
  _computing_data_collection
  _computing_data_reduction
  _computing_molecular_graphics
  _computing_publication_material
  _computing_structure_refinement
  _computing_structure_solution

The items in this category identify the software packages used for particular stages in a standard small-molecule crystallographic study. They may of course be used in other types of study as long as the description implied by the data name is relevant. The mmCIF dictionary provides a more general category, SOFTWARE, for the structured recording of programs used for a wider range of purposes.

3.2.5.3. Citation of related database entries

The single category describing related database entries is as follows:

DATABASE group
  DATABASE

Data items in this category are as follows:

DATABASE
  _database_code_CAS
  _database_code_CSD
  _database_code_ICSD
  _database_code_MDF
  _database_code_NBS
  _database_code_PDB
  _database_code_PDF
  _database_code_depnum_ccdc_archive
  _database_code_depnum_ccdc_fiz
  _database_code_depnum_ccdc_journal
  _database_CSD_history
  _database_journal_ASTM
  _database_journal_CSD

The _database_code_ items store the identifiers provided by specific databases for the structure described in the current data block. In the order given above, the databases they refer to are: *Chemical Abstracts*, the Cambridge Structural Database, the Inorganic Crystal Structure Database, the Metals Data File, the Crystal Data database of the National Institute of Standards and Technology (formerly the National Bureau of Standards), the Protein Data Bank, and the Powder Diffraction File of the International Centre for Diffraction Data.

The _database_code_depnum_ccdc_* items record deposition numbers assigned to files containing structural information archived by the Cambridge Crystallographic Data Centre (CCDC). The deposition numbers are as assigned by the CCDC itself (*_archive), by the Fachinformationszentrum Karlsruhe (*_fiz) or by a journal (*_journal). The item _database_CSD_history records the history of changes made by the CCDC and incorporated into the Cambridge Structural Database.

The _database_journal_ items store, respectively, the coden designator for journal titles of the American Society for Testing and Materials (ASTM), as given in the *Chemical Source List* maintained by the *Chemical Abstracts* Service, and the journal code used in the Cambridge Structural Database.

These specific items are regarded as appropriate for small-molecule and inorganic structures. The mmCIF dictionary includes a DATABASE_2 category, where an extensible data scheme allows additional database entries to be stored without requiring a separate data item for each new database reference.

3.2.5.4. Journal housekeeping, citation and indexing entries

The categories used for journal housekeeping and indexing are as follows:

JOURNAL group
  JOURNAL
  JOURNAL_INDEX

The data items in the JOURNAL category group are concerned with the processing of an article for publication. They are used mainly by the staff of the editorial office of an academic journal and are of limited interest to the practising crystallographer. They are not defined explicitly in the core dictionary and are included here only for the sake of completeness.

(*a*) JOURNAL
  _journal_coden_ASTM
  _journal_coden_Cambridge
  _journal_coeditor_address
  _journal_coeditor_code
  _journal_coeditor_email
  _journal_coeditor_fax
  _journal_coeditor_name
  _journal_coeditor_notes
  _journal_coeditor_phone
  _journal_data_validation_number
  _journal_date_accepted
  _journal_date_from_coeditor
  _journal_date_to_coeditor
  _journal_date_printers_final
  _journal_date_printers_first
  _journal_date_proofs_in
  _journal_date_proofs_out
  _journal_date_recd_copyright
  _journal_date_recd_electronic
  _journal_date_recd_hard_copy
  _journal_issue
  _journal_language
  _journal_name_full
  _journal_page_first
  _journal_page_last
  _journal_paper_category
  _journal_suppl_publ_number
  _journal_suppl_publ_pages
  _journal_techeditor_address
  _journal_techeditor_code
  _journal_techeditor_email
  _journal_techeditor_fax
  _journal_techeditor_name
  _journal_techeditor_notes
  _journal_techeditor_phone
  _journal_volume
  _journal_year

(*b*) JOURNAL_INDEX
  _journal_index_subterm
  _journal_index_term
  _journal_index_type

Of the data items in the JOURNAL category, the only ones that are likely to be of interest to users other than the journal staff are the items recording the bibliographic information upon publication, namely _journal_name_full, _journal_year, _journal_volume, _journal_page_first and _journal_page_last.

Data items in the JOURNAL_INDEX category allow terms to be embedded within a CIF that will be used for generating journal indexes. Example 3.2.5.2 shows how this is done; the possible values of _journal_index_type are defined by the journal and for *Acta Crystallographica* and other IUCr journals may be one of S (subject index), I (inorganic formula index), M (metal-organic formula index) or O (organic formula index).

Example 3.2.5.2. *Markup of indexing terms.*

```
loop_
     _journal_index_type
     _journal_index_term
     _journal_index_subterm
      O   C16H19NO4             .
      S   alkaloids            (-)-norcocaine
      S   (-)-norcocaine        .
      S
; [2R,3S-(2,3)]-methyl
  3-(benzoyloxy)-8-azabicyclo[3.2.1]octane-2-
  carboxylate
;                          .
```

Example 3.2.5.3. *Request to add material for publication to a journal's standard list.*

```
loop_
     _publ_manuscript_incl_extra_item
     _publ_manuscript_incl_extra_info
'_atom_site_symmetry_multiplicity'
     'to emphasise special sites'
'_chemical_compound_source'
     'rare material, unusual source'
'_reflns_d_resolution_high'
     'limited data are a problem here'
```

3.2.5.5. Contents of a publication

Categories used to describe an article for publication and to include the text of an article are as follows:

PUBL group

- PUBL
- PUBL_AUTHOR
- PUBL_BODY
- PUBL_MANUSCRIPT_INCL

The items in the PUBL category group describe the text that an author adds to the experimental data in a CIF to create a full record of the structural study for publication.

Data items in these categories are as follows:

(*a*) PUBL

```
_publ_contact_author
_publ_contact_author_address
_publ_contact_author_email
_publ_contact_author_fax
_publ_contact_author_id_iucr
_publ_contact_author_name
_publ_contact_author_phone
_publ_contact_letter
_publ_manuscript_creation
_publ_manuscript_processed
_publ_manuscript_text
_publ_requested_category
_publ_requested_coeditor_name
_publ_requested_journal
_publ_section_abstract
_publ_section_acknowledgements
_publ_section_comment
_publ_section_discussion
_publ_section_experimental
_publ_section_exptl_prep
_publ_section_exptl_refinement
_publ_section_exptl_solution
_publ_section_figure_captions
_publ_section_introduction
_publ_section_references
_publ_section_synopsis
_publ_section_table_legends
_publ_section_title
_publ_section_title_footnote
```

(*b*) PUBL_AUTHOR

```
_publ_author_address
_publ_author_email
_publ_author_footnote
_publ_author_id_iucr
_publ_author_name
```

(*c*) PUBL_BODY

```
_publ_body_contents
_publ_body_element
_publ_body_format
_publ_body_label
_publ_body_title
```

(*d*) PUBL_MANUSCRIPT_INCL

```
_publ_manuscript_incl_extra_defn
_publ_manuscript_incl_extra_info
_publ_manuscript_incl_extra_item
```

The data items in the PUBL category represent non-looped components of the published article, varying from the article title to the complete text of the article. Some journals such as *Acta Crystallographica* require specific section headers in articles, for which data items (*e.g.* `_publ_section_comment`) are provided. An alternative approach is to use the general items in this list for the article title, abstract, reference list *etc.* and build the individual sections of text using the items in the PUBL_BODY category.

The CIF syntax restrictions that permit only printable ASCII characters (Chapter 2.2) mean that authors cannot simply cut and paste text produced by commercial word-processing programs into a CIF. This might be inconvenient for the author, but while commercial word-processing programs are often convenient to use, they use proprietary and often poorly documented formats. For an archived CIF to remain readable in the long term, the use of transparent text representations, using open and well documented markup systems such as XML or TEX, is preferred.

The authors of an article are listed separately using items in the PUBL_AUTHOR category. The entry for each author can be annotated, for example to add text that would appear as a footnote to the author's name in the published article.

The PUBL_BODY category allows the body of an article to be more highly structured than `_publ_manuscript_text` does. It may be used for articles that include structural data but are less formally structured than required by *Acta Crystallographica Section C* or *Acta Crystallographica Section E*.

Journals like *Acta Crystallographica Section C* may have a list of CIF data items that will normally be published. If an author wishes to include additional data items, they can be specified using the _PUBL_MANUSCRIPT_INCL category. Since the *values* of `_publ_manuscript_incl_extra_item` are data names, they *must* be placed in quotes, as in Example 3.2.5.3, for them to be parsed correctly.

Further information on the use of the data items in the PUBL category group may be found in Section 5.7.2.

3.2.6. File metadata

The categories describing the history of a data block and its relation to other blocks are as follows:

AUDIT group

- AUDIT
- AUDIT_AUTHOR
- AUDIT_CONFORM
- AUDIT_CONTACT_AUTHOR
- AUDIT_LINK

Information about the origin and purpose of a CIF is needed to be able to make full use of the content of the CIF. Information about the CIF itself (rather than the experiment or structural model it describes) is known as *metadata*.

Because the scope of any data value is restricted to the data block in which it resides, each data block should contain its own set of **_audit_*** data items (a requirement that is often overlooked in the construction of a CIF with multiple data blocks). The data items in the AUDIT_LINK category may be used to record relationships between different data blocks within the same file.

Data items in these categories are as follows:

(*a*) AUDIT
```
_audit_block_code
_audit_creation_date
_audit_creation_method
_audit_update_record
```

(*b*) AUDIT_AUTHOR
```
_audit_author_address
_audit_author_name
```

(*c*) AUDIT_CONFORM
```
_audit_conform_dict_location
_audit_conform_dict_name
_audit_conform_dict_version
```

(*d*) AUDIT_CONTACT_AUTHOR
```
_audit_contact_author_address
_audit_contact_author_email
_audit_contact_author_fax
_audit_contact_author_name
_audit_contact_author_phone
```

(*e*) AUDIT_LINK
```
_audit_link_block_code
_audit_link_block_description
```

The AUDIT category provides a small set of data names suitable for identifying a data block and recording its creation date and subsequent modifications. Each data block in a CIF is introduced by a string of the form **data_*xxxx***, where the block code ***xxxx*** is an arbitrary string. CIF offers no guidelines for choosing a block code, and there are many cases where the same string has been chosen to label data blocks in different files. The **_audit_block_code** data item is meant to encourage authors to provide a unique label for a data block. Also, as a separate data item, **_audit_block_code** has the advantage that it can be interrogated using standard CIF query tools; this is not true of the block code.

The core dictionary does not specify a procedure for choosing a unique identifier for the data block, but other dictionaries do. The modulated structures dictionary recommends specific naming procedures (Section 3.4.4.4) and the power dictionary supplies alternative data items designed to generate globally unique identifiers (Section 3.3.7.1).

Some applications modify the block code in the **data_*xxxx*** string. The value of **_audit_block_code** may not be changed arbitrarily to suit the convenience of external applications.

In Example 3.2.6.1, the **_audit_block_code** assigned is different from the data-block code; the creation date is expressed in the CIF date format convention of *yyyy-mm-dd* and the revision record is generated by adding material to the **_audit_update_record** field. Each addition has been prefixed with the date and initialled by the person who made the change. It is good practice to maintain a full record of any changes of substance to the contents of the data block.

Data items in the AUDIT_AUTHOR category record details of the author or authors of the data block. Where there is more than a single author, the names and addresses are looped. The use of these data items parallels that of the items in the PUBL_AUTHOR category; the difference is that the latter are used specifically to record details of authors of an article for publication. The AUDIT_AUTHOR category refers to the creators of a CIF data block regardless of its intended purpose.

Example 3.2.6.1. *Items identifying a data block and recording its revision history.*
```
data_example

_audit_block_code                 xyzzy_2002-04-05
_audit_creation_date              2002-04-05
_audit_creation_method            'SHELXL97'

_audit_update_record
; 2002-04-09  discussion added                       BM
  2002-04-17  coeditor number XY1234 assigned       BM
  2002-04-18  revised comment after referee report BM
;
```

Example 3.2.6.2. *The CIF dictionaries to which the data block conforms.*
```
loop_
  _audit_conform_dict_name
  _audit_conform_dict_version
  _audit_conform_dict_location
    cif_core.dic      2.3.1   .
    cif_pd.dic        1.0.1   .
    cif_local_my.dic  1.0
              /usr/local/dics/my_local_dictionary
```

Data items in the AUDIT_CONFORM category describe the version of the dictionary or dictionaries that contain the definitions of the data names in the current data block. It is very helpful to provide this information, so that applications software can locate the original definitions and validate the contents of the current data block against them (Example 3.2.6.2). The dictionary identifier **_audit_conform_dict_name** is essential. The version is less important, as the dictionaries are revised in such a way as to try to retain compatibility between versions, but may occasionally be useful if changes of substance have crept in between versions. The location specified by **_audit_conform_dict_location** is useful only for local applications; in general the public register of CIF dictionaries should be used to locate dictionary files (see Section 3.1.8.3).

Data items in the AUDIT_CONTACT_AUTHOR category record details of the name and address of the author to be contacted concerning the contents of the data block. The use of these data items parallels that of the items in the PUBL_CONTACT_AUTHOR category; the difference is that the latter are used specifically to record details of the contact author of an article for publication. The AUDIT_CONTACT_AUTHOR category refers to the creator of a CIF data block regardless of its intended purpose.

The original purpose of a CIF, to record the data relevant to a single-crystal structure determination, was quickly extended to include the creation of an article reporting several crystal structures, as well as to powder CIFs recording information about multiple phases, modulated-structure CIFs describing superimposed structures and macromolecular CIFs recording results of multiple refinement cycles. A mechanism is required to differentiate the purpose of an individual data block and its relationship to other data blocks in the same file. This is provided by the AUDIT_LINK category. Example 3.2.6.3 shows how a CIF of an article for publication might show the relationships between the data blocks in the file. Note that the link references the value of **_audit_block_code** in the referenced data block, *not* the data-block header string itself (although in this example, and in Example 3.2.6.4, they have the same value).

Example 3.2.6.3. *List of linked data blocks in a CIF.*

```
data_global

_audit_block_code       global

loop_
  _audit_link_block_code
  _audit_link_block_description
    .          'text of paper with two structures'
    morA_(1)   'structure 1 of 2'
    morA_(2)   'structure 2 of 2'
```

Example 3.2.6.4. *Complementary list of linked data blocks in a secondary block.*

```
data_morA_(1)

_audit_block_code       morA_(1)

loop_
  _audit_link_block_code
  _audit_link_block_description
    global     'text of paper with two structures'
    .          'structure 1 of 2'
    morA_(2)   'structure 2 of 2'
```

For many applications, it is enough for a statement of the links between the data blocks in a CIF to be included once only in the file, normally in the initial data block. However, for completeness and to permit consistency checking, it is best if the other data blocks in the file have complementary declarations (Example 3.2.6.4).

Current practice as described in the core dictionary restricts this reporting of links between data blocks to the contents of a single file. In principle, if `_audit_block_code` were known to have globally unique values in each distinct data block, the mechanism could be extended to permit inter-file linkage.

Appendix 3.2.1
Category structure of the core CIF dictionary

Table A3.2.1.1 provides an overview of the structure of the core CIF dictionary by informal category group and categories.

Table A3.2.1.1. *Categories in the core CIF dictionary*

Numbers in parentheses refer to the section of this chapter in which each category is described in detail.

- ATOM group (§3.2.4.1)
  - ATOM_SITE (§3.2.4.1.1)
  - ATOM_SITES (§3.2.4.1.2)
  - ATOM_TYPE (§3.2.4.1.3)
- AUDIT group (§3.2.6)
  - AUDIT (§3.2.6(*a*))
  - AUDIT_AUTHOR (§3.2.6(*b*))
  - AUDIT_CONFORM (§3.2.6(*c*))
  - AUDIT_CONTACT_AUTHOR (§3.2.6(*d*))
  - AUDIT_LINK (§3.2.6(*e*))
- CELL group (§3.2.2.1)
  - CELL (§3.2.2.1(*a*))
  - CELL_MEASUREMENT_REFLN (§3.2.2.1(*b*))
- CHEMICAL group (§3.2.4.2)
  - CHEMICAL (§3.2.4.2.1(*a*))
  - CHEMICAL_CONN_ATOM (§3.2.4.2.2(*a*))
  - CHEMICAL_CONN_BOND (§3.2.4.2.2(*b*))
  - CHEMICAL_FORMULA (§3.2.4.2.1(*b*))
- CITATION group (§3.2.5.1)
  - CITATION (§3.2.5.1(*a*))
  - CITATION_AUTHOR (§3.2.5.1(*b*))
  - CITATION_EDITOR (§3.2.5.1(*c*))
- COMPUTING group (§3.2.5.2)
  - COMPUTING (§3.2.5.2)
- DATABASE group (§3.2.5.3)
  - DATABASE (§3.2.5.3)
- DIFFRN group (§3.2.2.2)
  - DIFFRN (§3.2.2.2.1)
  - DIFFRN_ATTENUATOR (§3.2.2.2.2(*a*))
  - DIFFRN_DETECTOR (§3.2.2.2.4)
  - DIFFRN_MEASUREMENT (§3.2.2.2.3(*a*))
  - DIFFRN_ORIENT_MATRIX (§3.2.2.2.3(*b*))
  - DIFFRN_ORIENT_REFLN (§3.2.2.2.3(*c*))
  - DIFFRN_RADIATION (§3.2.2.2.2(*b*))
  - DIFFRN_RADIATION_WAVELENGTH (§3.2.2.2.2(*c*))
  - DIFFRN_REFLN (§3.2.2.2.5(*a*))
  - DIFFRN_REFLNS (§3.2.2.2.5(*b*))
  - DIFFRN_REFLNS_CLASS (§3.2.2.2.5(*c*))
  - DIFFRN_SCALE_GROUP (§3.2.2.2.5(*d*))
  - DIFFRN_SOURCE (§3.2.2.2.2(*d*))
  - DIFFRN_STANDARD_REFLN (§3.2.2.2.5(*e*))
  - DIFFRN_STANDARDS (§3.2.2.2.5(*f*))
- EXPTL group (§3.2.2.3)
  - EXPTL (§3.2.2.3(*a*))
  - EXPTL_CRYSTAL (§3.2.2.3(*b*))
  - EXPTL_CRYSTAL_FACE (§3.2.2.3(*c*))
- GEOM group (§3.2.4.3)
  - GEOM (§3.2.4.3.1(*a*))
  - GEOM_ANGLE (§3.2.4.3.1(*b*))
  - GEOM_BOND (§3.2.4.3.1(*c*))
  - GEOM_CONTACT (§3.2.4.3.1(*d*))
  - GEOM_HBOND (§3.2.4.3.1(*e*))
  - GEOM_TORSION (§3.2.4.3.1(*f*))
- JOURNAL group (§3.2.5.4)
  - JOURNAL (§3.2.5.4(*a*))
  - JOURNAL_INDEX (§3.2.5.4(*b*))
- PUBL group (§3.2.5.5)
  - PUBL (§3.2.5.5(*a*))
  - PUBL_AUTHOR (§3.2.5.5(*b*))
  - PUBL_BODY (§3.2.5.5(*c*))
  - PUBL_MANUSCRIPT_INCL (§3.2.5.5(*d*))
- REFINE group (§3.2.3.1)
  - REFINE (§3.2.3.1(*a*))
  - REFINE_LS_CLASS (§3.2.3.1(*b*))
- REFLN group (§3.2.3.2)
  - REFLN (§3.2.3.2.1)
  - REFLNS (§3.2.3.2.2(*a*))
  - REFLNS_CLASS (§3.2.3.2.2(*b*))
  - REFLNS_SCALE (§3.2.3.2.2(*c*))
  - REFLNS_SHELL (§3.2.3.2.2(*d*))
- SYMMETRY group (§3.2.4.4)
  - SPACE_GROUP (§3.2.4.4.2(*a*))
  - SPACE_GROUP_SYMOP (§3.2.4.4.2(*b*))
  - SYMMETRY (§3.2.4.4.1(*a*))
  - SYMMETRY_EQUIV (§3.2.4.4.1(*b*))
- VALENCE group (§3.2.4.5)
  - VALENCE_PARAM (§3.2.4.5(*a*))
  - VALENCE_REF (§3.2.4.5(*b*))

References

Busing, W. R. & Levy, H. A. (1967). *Angle calculations for 3- and 4-circle X-ray and neutron diffractometers. Acta Cryst.* **22**, 457–464.

Hall, S. R. (1981). *Space-group notation with an explicit origin. Acta Cryst.* A**37**, 517–525; erratum (1981), A**37**, 921.

Hall, S. R. & Grosse-Kunstleve, R. W. (2001). *International tables for crystallography*, Vol. B, *Reciprocal space*, edited by U. Shmueli, 2nd ed., Appendix A1.4.2.3. Dordrecht: Kluwer Academic Publishers.

Klyne, W. & Prelog, V. (1960). *Description of steric relationships across single bonds. Experientia*, **16**, 521–523.