3.6.7.6. Crystal symmetry

The categories describing symmetry are as follows:

SYMMETRY group

SYMMETRY

SYMMETRY_EQUIV

SPACE_GROUP

SPACE_GROUP_SYMOP

Data items in the SYMMETRY category are used to give details about the crystallographic symmetry. The equivalent positions for the space group are listed using data items in the SYMMETRY_EQUIV category. These categories are used in the same way in the core CIF and mmCIF dictionaries, and Section 3.2.4.4 can be consulted for details.

The current version of the mmCIF dictionary includes the SPACE_GROUP categories that were derived from the symmetry CIF dictionary (Chapter 3.8) and included in version 2.3 of the core CIF dictionary. At the time of writing, macromolecular applications have not yet begun to make use of these new categories.

Data items in these categories are as follows:

(*a*) SYMMETRY

- **_symmetry.entry_id**
  → **_entry.id**
  _symmetry.cell_setting
  _symmetry.Int_Tables_number
  _symmetry.space_group_name_Hall
  _symmetry.space_group_name_H-M

(*b*) SYMMETRY_EQUIV

- **_symmetry_equiv.id** (∼ **_symmetry_equiv_pos_site_id**)
  _symmetry_equiv.pos_as_xyz

(*c*) SPACE_GROUP

- **_space_group.id**
  _space_group.crystal_system
  _space_group.IT_number
  _space_group.name_H-M_alt
  _space_group.name_Hall

(*d*) SPACE_GROUP_SYMOP

- **_space_group_symop.id**
  _space_group_symop.operation_xyz
  _space_group_symop.sg_id

The bullet (•) *indicates a category key. The arrow* (→) *is a reference to a parent data item. Items in italics have aliases in the core CIF dictionary formed by changing the full stop* (.) *to an underscore* (_) *except where indicated by the* ∼ *symbol.*

The data item **_symmetry.entry_id** has been added to the SYMMETRY category to provide the formal category key required by the DDL2 data model.

3.6.7.7. Bond-valence information

The categories describing bond valences are as follows:

VALENCE group

VALENCE_PARAM

VALENCE_REF

These categories were introduced into version 2.2 of the core CIF dictionary to provide the information about bond valences required in inorganic crystallography. They appear in the mmCIF dictionary only for full compatibility with the core dictionary.

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 bullet (•) *indicates a category key. The arrow* (→) *is a reference to a parent data item. Items in italics have aliases in the core CIF dictionary formed by changing the full stop* (.) *to an underscore* (_).

Information about the use of these data items in the core CIF dictionary is given in Section 3.2.4.5.

3.6.8. Publication

The results of the determination of the crystal structure of a biological macromolecule might be published in an academic journal and/or deposited in a structural database. The data items in the core CIF dictionary cover most of the requirements for constructing an article for publication from an mmCIF and the many well defined data fields in mmCIF allow an extensively annotated record of the structure to be deposited in a database. However, the formalism of two of the core CIF categories for publication did not fit the relational database model of mmCIF, so new categories were required. The core CIF category COMPUTING, which is used to list the programs used to determine the structure, is replaced by the mmCIF category SOFTWARE, and the core CIF category DATABASE, which is used to identify the records associated with the structure in various databases, is replaced by the mmCIF category DATABASE_2.

The category groups discussed here are: the CITATION group, which is used to give citations to the literature (Section 3.6.8.1); the COMPUTING group, which is used to cite software (Section 3.6.8.2); the DATABASE group for citing related database entries (Section 3.6.8.3), which includes a group of categories used to ensure compatibility with specific database records in the Protein Data Bank (Section 3.6.8.3.2); journal administration categories that might be used by a publisher (Section 3.6.8.4.1); and the PUBL family of categories used to store the text of an article for publication (Section 3.6.8.4.2).

3.6.8.1. Literature citations

The categories describing literature citations are as follows:

CITATION group

CITATION

CITATION_AUTHOR

CITATION_EDITOR

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.details (~ _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. Items in italics have aliases in the core CIF dictionary formed by changing the full stop (.) to an underscore (_).

The original core CIF dictionary contained the data item **_publ_section_references** for citations of journal articles, book chapters and monographs. The authors of the mmCIF dictionary felt that a more detailed and structured approach to literature citations was required. This is provided by the mmCIF categories CITATION, CITATION_AUTHOR and CITATION_EDITOR. These categories were subsequently included in the core CIF dictionary and are used in the same way in both dictionaries. Section 3.2.5.1 may be consulted for details. Although **_publ.section_references** remains a valid mmCIF data item, it is expected that the CITATION, CITATION_AUTHOR and CITATION_EDITOR categories will be used for literature citations in mmCIFs.

3.6.8.2. Citation of software packages

The categories describing software citations are as follows:

COMPUTING group

COMPUTING

SOFTWARE

It is expected that citations of software packages in an mmCIF will be made using data items in the SOFTWARE category. However, in some cases, a particular publisher or database may require that this information is given using data items in the COMPUTING category instead (see Section 3.2.5.2 for details).

Data items in these categories are as follows:

(*a*) COMPUTING

```
• _computing.entry_id
        → _entry.id
  _computing.cell_refinement
  _computing.data_collection
  _computing.data_reduction
  _computing.molecular_graphics
  _computing.publication_material
  _computing.structure_refinement
  _computing.structure_solution
```

(*b*) SOFTWARE

```
• _software.name
• _software.version
  _software.citation_id
        → _citation.id
  _software.classification
  _software.compiler_name
  _software.compiler_version
  _software.contact_author
  _software.contact_author_email
  _software.date
  _software.dependencies
  _software.description
  _software.hardware
  _software.language
  _software.location
  _software.mods
  _software.os
  _software.os_version
  _software.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. Items in italics have aliases in the core CIF dictionary formed by changing the full stop (.) to an underscore (_).

The data item **_computing.entry_id** has been added to the COMPUTING category to provide the formal category key required by the DDL2 data model.

The data items in the SOFTWARE category are used to cite the software packages used in the structure analysis. The software can be described in great detail if necessary. However, for most applications a small subset of these data items, for example just **_software.name** and **_software.version**, could be used (see Example 3.6.8.1).

Most data items in the SOFTWARE category are self-explanatory, but a few require further comment. The data item **_software.citation_id** provides a way to link the details of a program to the citation of an article in the literature that describes the program; this data item must match a value of **_citation.id** in the CITATION category. The name and e-mail address of the author of the software can also be given using **_software.contact_author** and **_software.contact_author_email**, respectively. (This may be the original author or someone who subsequently modifies or maintains the software; these data items would generally refer to the person most closely associated with the maintenance of the code at the time it was used.) The release date of the software may be recorded in **_software.date**. As far as possible, the date should be that of the version recorded in **_software.version**. The data item **_software.location** may be used to supply a URL from which the software may be downloaded or where it is described in detail.

3.6.8.3. Citation of related database entries

Categories describing related database entries are as follows:

DATABASE group

Related database entries (§3.6.8.3.1)

DATABASE

DATABASE_2

Example 3.6.8.1. *The refinement program Prolsq described with data items in the SOFTWARE category.*

```
_software.name
_software.version
_software.date
_software.type
_software.contact_author
_software.contact_author_email
_software.location
_software.classification
_software.citation_id
_software.language
_software.compiler_name
_software.compiler_version
_software.hardware
_software.os
_software.os_version
_software.dependencies
_software.mods
_software.description
   Prolsq    unknown  .    program
    'Wayne A. Hendrickson'  ?
    'ftp://rosebud.sdsc.edu/pub/sdsc/xtal/CCP4/ccp4/'
    refinement ref5  Fortran
  'Convex Fortran' v8.0 'Convex C220' ConvexOS  v10.1
    'Requires that Protin be run first'    optimized
    'restrained least-squares refinement'
```