

## 3. CIF DATA DEFINITION AND CLASSIFICATION

```

data_mmcif_std.dic

_dictionary.title          mmcif_std.dic
_dictionary.version       2.0.09
_dictionary.datablock_id  mmcif_std.dic
                          (a)

loop_
_dictionary_history.version
_dictionary_history.update
_dictionary_history.revision . . .
                          (b)

loop_
_sub_category.id
_sub_category.description . . .

loop_
_category_group_list.id
_category_group_list.parent_id
_category_group_list.description . . .
                          (c)

loop_
_item_type_list.code
_item_type_list.primitive_code
_item_type_list.construct
_item_type_list.detail

loop_
_item_units_list.code
_item_units_list.detail . . .

loop_
_item_units_conversion.from_code
_item_units_conversion.to_code
_item_units_conversion.operator
_item_units_conversion.factor . . . .
                          (d)

save_CATEGORY_A . . . save_
save_category_a.item_1 . . . save_
save_category_a.item_2 . . . save_
save_category_a.item_3 . . . save_

save_CATEGORY_B . . . save_
save_category_b.item_1 . . . save_
save_category_b.item_2 . . . save_
                          (e)

```

Fig. 3.1.6.1. Schematic structure of the macromolecular CIF dictionary. (a) Dictionary identifiers. (b) Dictionary history. (c) Subcategory and category group listings. (d) Data types, units descriptions and conversion tables. (e) Multiple category and item definition blocks.

items are defined in the same data block and are understood to share the common attributes itemized in that data block).

Within DDL2, there are mechanisms for more formal and machine-parsable statements of relationships. The `_sub_category.id` attribute is a label shared by several data items within a category that are related in a specific way described by the associated `_sub_category.description` attribute. The relationships may be rather general, such as elements of a matrix; or they may be specific physical properties or attributes, such as the collection of axis lengths of a unit cell. The dictionary should list all such labels that occur within its included data definition blocks. Example 3.1.6.2 is an extract from the macromolecular dictionary.

### 3.1.6.3. Category groupings

In the DDL2 data model, a *category* of data corresponds to a set of related data items that may be stored in a single relational

#### Example 3.1.6.1. DDL2 dictionary identification entries.

```

data_mmcif_std.dic

_dictionary.title          mmcif_std.dic
_dictionary.version       2.0.09
_dictionary.datablock_id  mmcif_std.dic

loop_
_dictionary_history.version
_dictionary_history.update
_dictionary_history.revision
0.1.1 1993-02-11
; Highlighted all notes with # %%%% surrounds.
;
. . .

```

#### Example 3.1.6.2. DDL2 subcategories defined in the mmCIF dictionary.

```

loop_
_sub_category.id
_sub_category.description
'fractional_coordinate'
; The collection of x, y, and z components of a
position specified with reference to unit cell
directions.
;
'matrix'
; The collection of elements of a matrix.
;
'miller_index'
; The collection of h, k, and l components of the
Miller index of a reflection.
;
'cell_length'
; The collection of a, b, and c axis lengths of a
unit cell.
;
'mm_atom_site_label'
; The collection of alt id, asym id, atom id, comp id
and seq id components of the label for a
macromolecular atom site.
;

```

database table. A number of such tables may collectively describe the complete properties of some physical object. This is expressed formally by assigning the same label (`_category_group.id`) to the relevant categories. While relationships between categories are implied in DDL1 dictionaries by the hierarchical structure of the names of data items, in DDL2 dictionaries the relationships are formally stated.

For subcategories, the category-group relationships present in the dictionary are listed in a separate looped list. Example 3.1.6.3 is an extract from the macromolecular dictionary. The `inclusive_group` entry shows the common parentage of all categories (and ultimately all data items) in the dictionary.

### 3.1.6.4. Category definitions

In the DDL2 formalism, a category of data items may be mapped to a relational table. The dictionary entry for a category includes the name of the category (an identifying label which is referenced by the `_item.category_id` attribute of each component data item) and a list of the category groups of which it may be considered a member. The category *key* is explicitly specified – that is, the data item (or group of items) that uniquely identifies an individual row in a table of data of that category.

Where a category encompasses a set of data items that are not normally specified in a looped list, the category may nevertheless be taken to represent a degenerate table with a single row, and therefore there is still a category key. For degenerate categories the key value is often set equal to the name of the parent data block.