1 Dictionary Structure

1.1 Dictionary Definition Language

The framework for mmCIF dictionary is defined by the Dictionary Description Language (DDL). The role of the DDL is to define the data items which may be used to construct each definitions in the mmCIF dictionary, and also to define the relationships between these defining data items. The DDL is expressed as a dictionary using its own definitional content. The text form of the current version of the DDL dictionary is here.

The DDL contains no information about macromolecular structure, rather it defines data items which can be used to describe other data. The DDL is actually quite generic. It defines data items that describe the general features of a data item like a textual description, a data type, a set of examples, a range of permissible values, or perhaps a discrete set of permitted values.

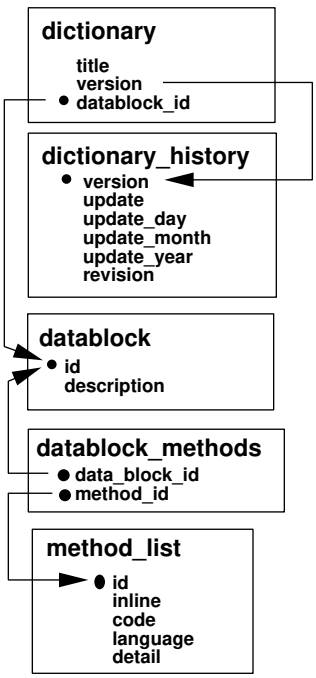
The DDL combines collections of related data items into categories. A categories is essentially a table in which each repetition of the group of related items adds a row. Within a category, those data items which determine the uniqueness of their group are designated as key items in the category. No data item group in a category is allowed to have a set of duplicate values of its key items. Each data item is assigned membership in one or more categories. Parent-child relationships may be specified for items which belong to multiple categories. These relationships permit the expression of the very complicated data structures required to describe macromolecular structure.

The DDL also provides some other levels of data organization in addition to the category. Related categories may be collected together in category groups, and parent relationships may be specified for these groups. This higher level of association provides a means of organizing large complicated collection of categories into smaller more relevant and potentially interrelated groups. Within the level of a category, subcategories of data items may be defined among groups of related data items. The subcategory provides a mechanism to identify that, for example, the data items month, day, and year collectively define a date.

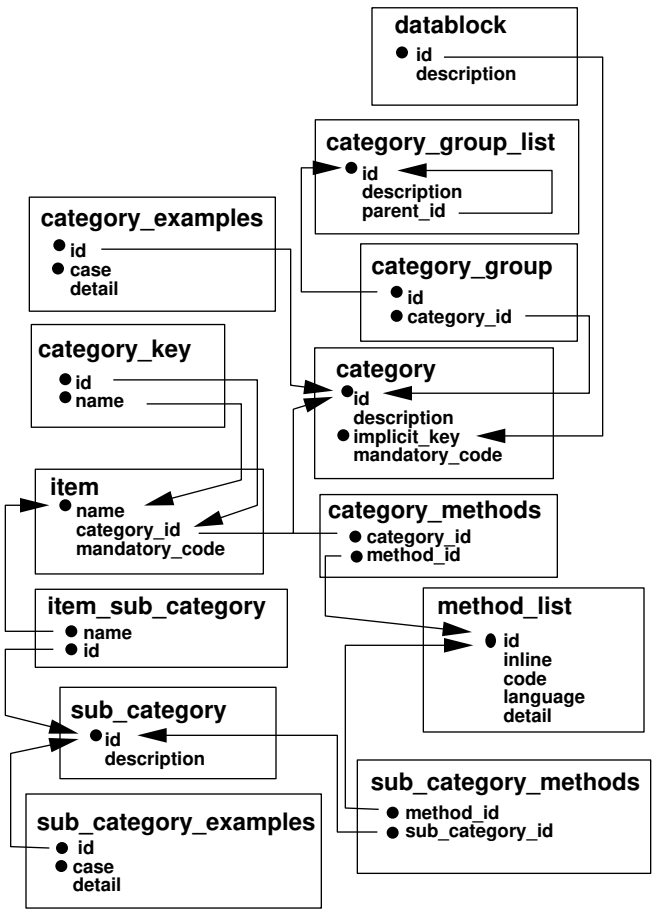
The highest levels of data organization provided by the DDL are the data block and the dictionary. The dictionary level collects a set of related definitions into a single unit, and provides for a detailed revision history to be maintained on the collection. The data block level ties the contents of a dictionary to the `data_` section in which it is contained. The identifier for the data block and hence the dictionary is added implicitly to the key of each category.

The following sections provide schematic diagrams of each of the organizational features provided by the DDL. In these diagrams, boxes enclose the the data items within each category. Key data items are preceded by a dark circle. Data items common to multiple categories are identified by connecting lines with the arrows pointing at the parent definition of the data item.

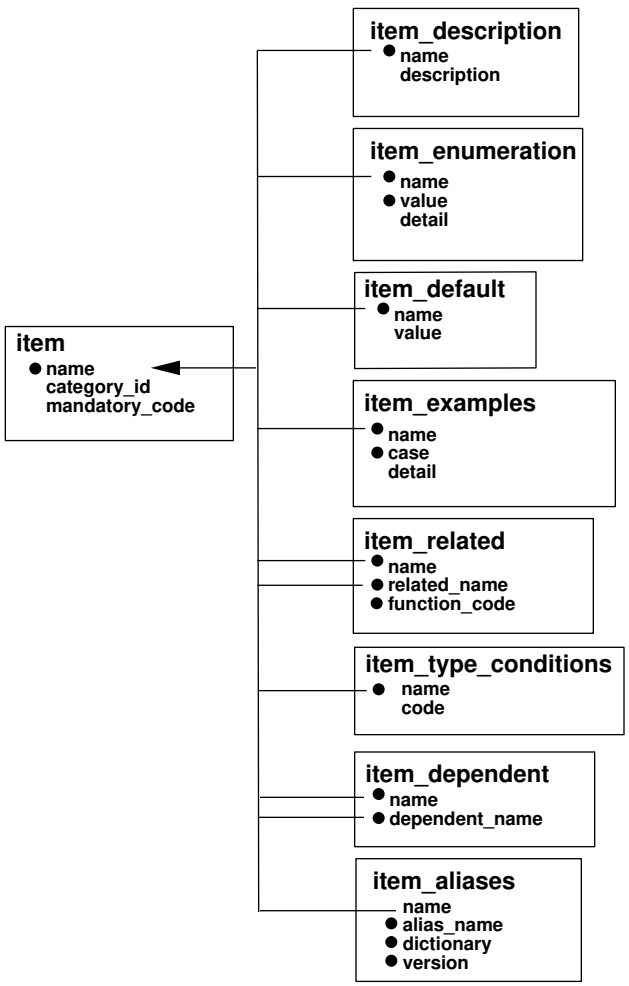
1.1.1 Dictionary Level of DDL Description

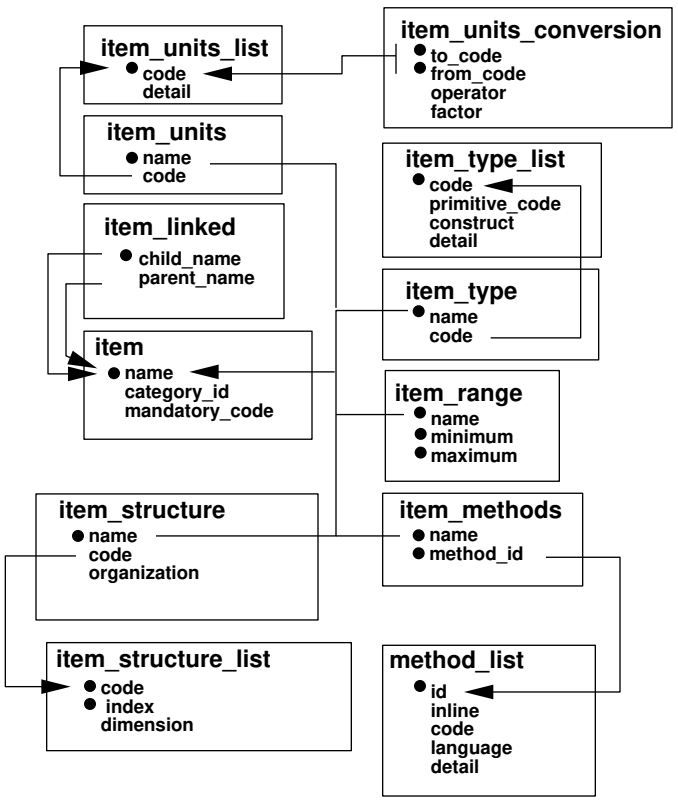


1.1.2 Category Level of DDL Description



1.1.3 Data Item Level of DDL Description


item_description
name
description
item_enumeration
name
value
detail
item_default
name
value
item
name
category_id
mandatory_code
item_examples
name
case
detail
item_related
name
related_name
function_code
item_type_conditions
name
code
item_dependent
name
dependent_name
item_aliases
name
alias_name
dictionary
version



1.2 Some Dictionary Examples

In this section, several examples are presented which illustrate how the elements of the DDL are combined into dictionary definitions.

1.2.1 An item definition: `_citation.journal_abbrev`

```
save__citation.journal_abbrev
     _item_description.description
;              Abbreviated name of the journal cited as given in the Chemical
               Abstracts Service Source Index.
;
     _item.name                  '_citation.journal_abbrev'
     _item.category_id             citation
     _item.mandatory_code          no
     _item_aliases.alias_name    '_citation_journal_abbrev'
     _item_aliases.dictionary      cif_core.dic
     _item_aliases.version         2.0.1
     _item_type.code               line
     _item_examples.case          'J. Mol. Biol.'
     save_
```

1.2.1.1 Category ITEM_DESCRIPTION

The ITEM_DESCRIPTION category holds a text description of each data item which is typically written as a definition of the item.

1.2.1.2 Category ITEM

The ITEM category holds the item name, category name and a code indicating if this item mandatory in any row of this category. The value of the mandatory code is either yes, no, or implicit. The implicit value is used to indicate that a value is required for the item but it can be derived from the context of the definition and need not be specified. This feature is most often used in the DDL to indicate that item name values can be derived from the name of the save frame in which they are defined.

Note that the value of the _item.name in the above example is enclosed in quotation marks. This is a requirement of the syntax and avoids confusing data values with item names.

1.2.1.3 Category ITEM_ALIASES

The mmCIF dictionary contains a superset of the definitions that were originally defined in the CIF core dictionary cif_core.dic. In order to maintain backward compatibility with original definitions, the ITEM_ALIASES category was introduced to hold the item name, dictionary name and version in which the original definition of an item was published.

1.2.1.4 Category ITEM_TYPE

The ITEM_TYPE category holds a reference to a data type defined in the ITEM_TYPE_LIST category. A reference to the data type is used here rather that a detailed data type description in order to avoid repeating such description for other data items. A single list of data types and associated regular expressions is created in the ITEM_TYPE_LIST and this referenced by all of the definitions in a dictionary. In the mmCIF dictionary the codes that are used to described the data types are generally easy to interpret. In this case, a single line of text will be accepted for this data item.

1.2.1.5 Category ITEM_EXAMPLES

Examples of a data item can be included in the ITEM_EXAMPLES category. In this case only a single example has been provided, but many examples can be provided by using a loop_ directive.

1.2.2 Related item definitions: _cell.length_a and _cell.length_a_esd

```
save__cell.length_a
    _item_description.description
;               Unit-cell length a corresponding to the structure reported.
;
    _item.name                  '_cell.length_a'
    _item.category_id             cell
    _item.mandatory_code          no
    _item_aliases.alias_name    '_cell_length_a'
    _item_aliases.dictionary      cif_core.dic
    _item_aliases.version         2.0.1
```

```
     loop_
    _item_dependent.dependent_name
                                ’_cell.length_b’
                                ’_cell.length_c’
     loop_
    _item_range.maximum
    _item_range.minimum            .    0.0
                                 0.0   0.0
    _item_related.related_name  ’_cell.length_a_esd’
    _item_related.function_code   associated_esd
    _item_sub_category.id         cell_length
    _item_type.code               float
    _item_type_conditions.code    esd
    _item_units.code              angstroms
     save_

save__cell.length_a_esd
    _item_description.description
;               The estimated standard deviation of _cell.length_a.
;
    _item.name                  ’_cell.length_a_esd’
    _item.category_id             cell
    _item.mandatory_code          no
    _item_default.value           0.0
     loop_
    _item_dependent.dependent_name
                                ’_cell.length_b_esd’
                                ’_cell.length_c_esd’
    _item_related.related_name  ’_cell.length_a’
    _item_related.function_code   associated_value
    _item_sub_category.id         cell_length_esd
    _item_type.code               float
    _item_units.code              angstroms
     save_
```

1.2.2.1 Category `ITEM_DEPENDENT`

Some data items are only meaningful when expressed as a complete set. The `ITEM_DEPENDENT` category is used to store this type of information. Those additional data items within the category which are required for the meaningful interpretation of the a item are listed in this category. In the above example, the cell lengths in the `b` and `c` directions are defined as dependent items of the cell length in the `a` direction.

1.2.2.2 Category `ITEM_RANGE`

The permissible range of values for a numerical data item are stored in the `ITEM_RANGE` category. Each boundary condition is defined as the non-inclusive range between a pair of mininum and maximum values. If multiple boundary conditions are specified using the `loop_` directive, then each condition must be satisfied. A discrete boundary value may be set by assigning the desired value as both the maximum and minimum value. In the above example, the permissible cell length range is defined as greater than or equal to zero.

1.2.2.3 Category ITEM_RELATED

A number of special relationships may be defined between data items. For those relationships which occur frequently within the dictionary, the source or function of the relationship has been standardized. In the example above, this category is used to identify that the _cell.length_a_esd is the estimated standard deviation of _cell.length_a_esd. The recognized relationships are fully described in the DDL definition of the data item _item_related.function_code in catetory ITEM_RELATED. The current list includes:

alternate
: indicates that the item identified in _item_related.related_name is an alternative expression in terms of its application and attributes to the item in this definition.

alternate_exclusive
: indicates that the item identified in _item_related.related_name is an alternative expression in terms of its application and attributes to the item in this definition. Only one of the alternative forms may be specified.

convention
: indicates that the item identified in _item_related.related_name differs from the defined item only in terms of a convention in its expression.

conversion_constant
: indicates that the item identified in _item_related.related_name differs from the defined item only by a known constant.

conversion_arbitrary
: indicates that the item identified in _item_related.related_name differs from the defined item only by a arbitrary constant.

replaces
: indicates that the defined item replaces the item identified in _item_related.related_name.

replacedby
: indicates that the defined item is replaced by the item identified in _item_related.related_name.

associated_value
: indicates that the item identified in _item_related.related_name is meaningful when associated with the defined item.

associated_esd
: indicates that the item identified in _item_related.related_name is the estimated standard deviation of of the defined item.

1.2.2.4 Category ITEM_SUB_CATEGORY

Sets of data items within a category may be collected into named subcategories. ITEM_SUB_CATEGORY is used to assign subcategory membership to a data item. In the above example, item _cell.length_a is assigned to the subcategory CELL_LENGTH. Although not shown, items _cell.length_b and _cell.length_c are similarly added to this subcategory.

1.2.2.5 Category ITEM_UNITS

The ITEM_UNITS category holds the name of the system of units in which an item is expressed. The name assigned to _item_units.code refers to a single list of all of the unit types used in the dictionary. This list is stored in the category ITEM_UNITS_LIST. Conversion factors between different systems of units are provided in the data table stored in category ITEM_UNITS_CONVERSION.

1.2.3 A Category definition: CELL

```
save_CELL
    _category.description
;               Data items in the CELL category record details about the
                crystallographic cell parameters.
;
    _category.id                   cell
    _category.mandatory_code       no
    _category_key.name           '_cell.entry_id'
     loop_
    _category_group.id            'inclusive_group'
                                  'cell_group'
     loop_
    _category_examples.detail
    _category_examples.case
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
;
    Example 1 - based on PDB entry 5HVP and laboratory records for the
                structure corresponding to PDB entry 5HVP
;
;
    _cell.entry_id                          '5HVP'
    _cell.length_a                          58.39
    _cell.length_a_esd                       0.05
    _cell.length_b                          86.70
    _cell.length_b_esd                       0.12
    _cell.length_c                          46.27
    _cell.length_c_esd                       0.06
    _cell.angle_alpha                       90.00
    _cell.angle_beta                        90.00
    _cell.angle_gamma                       90.00
    _cell.volume                            234237
    _cell.details
    ; The cell parameters were refined every twenty frames during data
      integration. The cell lengths given are the mean of 55 such refinements;
      the esds given are the root mean square deviations of these 55
      observations from that mean.
    ;
;
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
;
    Example 2 - based on data set TOZ of Willis, Beckwith & Tozer [(1991).
                Acta Cryst. C47, 2276-2277].
;
;
    _cell.length_a                      5.959
    _cell.length_a_esd                  0.001
```

```
    _cell.length_b                    14.956
    _cell.length_b_esd                 0.001
    _cell.length_c                    19.737
    _cell.length_c_esd                 0.003
    _cell.angle_alpha                 90.0
    _cell.angle_beta                  90.0
    _cell.angle_gamma                 90.0
    _cell.volume                      1759.0
    _cell.volume_esd                     0.3
;
# - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
     save_
```

1.2.3.1 Category CATEGORY

The name and textual description of a category are stored in the category named CATEGORY. A code (_category.mandatory_code) to indicate if the category must appear in any data block based on this dictionary is also stored in this category.

1.2.3.2 Category CATEGORY_KEY

The list of data items which uniquely identify each row of a category are stored in the CATEGORY_KEY. In the example above, the item _cell.entry_id is defined as the category key. This item is a reference to the top level identifier in the mmCIF dictionary, _entry.id. Because only a single entry may exist within an mmCIF data block, this key assignment defines that only a single row may exist in the CELL category.

1.2.3.3 Category CATEGORY_GROUP

Membership in category groups is stored in the category CATEGORY_GROUP. Each category group in this category must have a corresponding definition in the category CATEGORY_GROUP_LIST. In the above example, the CELL category is assigned a category groups cell_group and inclusive_group. The former contains other categories which describe properties of the crystallographic cell, and the latter includes all of the categories in the mmCIF dictionary.

1.2.3.4 Category CATEGORY_EXAMPLES

Complete and annotated examples of a category are stored in the CATEGORY_EXAMPLES category. The text of the category example is stored in item _category_examples.case and any associated annotation is stored in item _category_examples.detail. Multiple examples are defined for the CELL category above.

1.2.4 Categories with a common item: CITATION and CITATION_AUTHOR

```
save_CITATION
    _category.description
;               Data items in the CITATION category record details about the
                literature cited relevant to the contents of the data block.
;
    _category.id                  citation
```

```
    _category.mandatory_code      no
    _category_key.name          '_citation.id'
     loop_
    _category_group.id           'inclusive_group'
                                 'citation_group'
#
#      ---------  Abbreviated Definition  ----------
     save_

save__citation.id
    _item_description.description
;               The value of _citation.id must uniquely identify a record in the
                CITATION list.

                The _citation.id 'primary' should be used to indicate the
                citation that the author(s) consider to be the most pertinent to
                the contents of the data block.

                Note that this item need not be a number; it can be any unique
                identifier.
;
     loop_
    _item.name
    _item.category_id
    _item.mandatory_code
                '_citation.id'                  citation         yes
                '_citation_author.citation_id'  citation_author  yes
                '_citation_editor.citation_id'  citation_editor  yes
                '_software.citation_id'         software         yes
    _item_aliases.alias_name    '_citation_id'
    _item_aliases.dictionary      cif_core.dic
    _item_aliases.version         2.0.1
     loop_
    _item_linked.child_name
    _item_linked.parent_name
                '_citation_author.citation_id'  '_citation.id'
                '_citation_editor.citation_id'  '_citation.id'
                '_software.citation_id'         '_citation.id'
    _item_type.code               code
     loop_
    _item_examples.case          'primary'
                                 '1'
                                 '2'
     save_


save_CITATION_AUTHOR
    _category.description
;               Data items in the CITATION_AUTHOR category record details
                about the authors associated with the citations in the
                CITATION list.
;
    _category.id                  citation_author
    _category.mandatory_code      no
     loop_
    _category_key.name          '_citation_author.citation_id'
                                '_citation_author.name'
```

```
     loop_
    _category_group.id           'inclusive_group'
                                 'citation_group'
#
#      ---------  Abbreviated Definition  ----------
     save_

save__citation_author.citation_id
    _item_description.description
;               This data item is a pointer to _citation.id in the CITATION
                category.
;
    _item.name                  '_citation_author.citation_id'
    _item.mandatory_code          yes
    _item_aliases.alias_name    '_citation_author_citation_id'
    _item_aliases.dictionary      cif_core.dic
    _item_aliases.version         2.0.1
     save_
```

1.2.4.1 Category ITEM

This example illustrates how an item which occurs in multiple categories is defined. In the above case of the citation identifier (_citation.id), the ITEM category is preceded by a loop_ directive and all of the definitions citation identifier are listed. For instance, the citation identifier is also an item in category CITATION_AUTHOR where it has the item name _citation_author.citation_id. For conformity with the manner in which the CIF core dictionary has been organized, a skeleton definition of the child data item _citation_author.citation_id has been included in the dictionary. In fact, this skeleton definition is formally unnecessary.

As a matter of style, the mmCIF dictionary generally defines all of the instances of a data item within the parent definition. Items which are related to the parent definition are listed in the ITEM_LINKED category.

1.2.4.2 Category ITEM_LINKED

The repitition of a data item in multiple categories gives rise to a parent-child relationships between such definitions. These relationships are stored in the ITEM_LINKED category. In the example above, this category stores the list of data items which are children of the citation identifier, _citation.id. These include _citation_author.citation_id, _citation_editor.citation_id and _software.citation_id.

Dictionary Structure

Table of Contents

1 Dictionary Structure 1
1.1 Dictionary Definition Language 1
1.1.1 Dictionary Level of DDL Description 1
1.1.2 Category Level of DDL Description 2
1.1.3 Data Item Level of DDL Description 3
1.2 Some Dictionary Examples 5
1.2.1 An item definition: _citation.journal_abbrev 5
1.2.1.1 Category ITEM_DESCRIPTION 6
1.2.1.2 Category ITEM 6
1.2.1.3 Category ITEM_ALIASES 6
1.2.1.4 Category ITEM_TYPE 6
1.2.1.5 Category ITEM_EXAMPLES 6
1.2.2 Related item definitions: _cell.length_a and _cell.length_a_esd . . 6
1.2.2.1 Category ITEM_DEPENDENT 7
1.2.2.2 Category ITEM_RANGE 7
1.2.2.3 Category ITEM_RELATED 8
1.2.2.4 Category ITEM_SUB_CATEGORY 8
1.2.2.5 Category ITEM_UNITS 9
1.2.3 A Category definition: CELL 9
1.2.3.1 Category CATEGORY 10
1.2.3.2 Category CATEGORY_KEY 10
1.2.3.3 Category CATEGORY_GROUP 10
1.2.3.4 Category CATEGORY_EXAMPLES 10
1.2.4 Categories with a common item: CITATION and CITATION_AUTHOR 10
1.2.4.1 Category ITEM 12
1.2.4.2 Category ITEM_LINKED 12