International Tables for Crystallography (2006). Vol. G, Section 3.6.8.4, pp. 193–194.

A major goal of the design of the mmCIF data model was that a file could be transformed from Protein Data Bank (PDB) format to mmCIF format and back again without loss of information. This required the creation of mmCIF data items whose sole purpose is to capture PDB-specific records that do not map onto mmCIF data items. These records would never be created for a *de novo* mmCIF. This family of categories also belongs to the PDB category group (see Section 3.6.9.3).

The items in the categories DATABASE_PDB_MATRIX and DATABASE_PDB_TVECT are derived from the elements of transformation matrices and vectors used by the Protein Data Bank. The items in the categories DATABASE_PDB_REV and DATABASE_PDB_REV_RECORD record details about the revision history of the data block as archived by the Protein Data Bank.

The items in the DATABASE_PDB_CAVEAT category record comments about the data block flagged as 'CAVEATS' by the Protein Data Bank at the time the original PDB archive file was created. A PDB CAVEAT record indicates that the entry contains severe errors. In PDB format, extended comments were stored as a sequence of fixed-length (80-character) format records, columns 9 and 10 being reserved for continuation sequence numbering. The mmCIF representation retains each record as a separate data value and does not attempt to merge continuation records to provide more readable running text. Hence the PDB CAVEAT entry

```
CAVEAT     1ABC    THE CRYSTAL TRANSFORMATION IS WRONG
CAVEAT   2 1ABC    BUT IS UNCORRECTABLE AT THIS TIME
```

would be represented in mmCIF as

```
loop_
  _database_PDB_caveat.id
     _database_PDB_caveat.text
    1
; THE CRYSTAL TRANSFORMATION IS WRONG
;
    2
; BUT IS UNCORRECTABLE AT THIS TIME
;
```

The PDB format used 'REMARK' records to store information relating to several aspects of the structure in free or loosely structured text. In some cases, the conventions used for individual types of REMARK record allow structured data to be extracted automatically and translated to specific mmCIF data items. Where this is not possible, the DATABASE_PDB_REMARK category may be used to retain the information that appeared in these parts of PDB format files. Unlike the CAVEAT records, it is possible to collect together several REMARK records sharing a common numbering into a single free-text field. For example, PDB practice has been to repeat the contents of CAVEAT records (see above) as records of type 'REMARK 5'. While each separate CAVEAT record is converted to a separate mmCIF data value, the complete text of a REMARK 5 record may be gathered into a single mmCIF data value. Hence the CAVEAT example above would also appear in a PDB file as part of a 'REMARK 5' as

```
REMARK   5 THE CRYSTAL TRANSFORMATION IS WRONG
REMARK   5 BUT IS UNCORRECTABLE AT THIS TIME
```

and would appear in an mmCIF as

```
loop_
  _database_PDB_remark.id
     _database_PDB_remark.text
    5
; THE CRYSTAL TRANSFORMATION IS WRONG
  BUT IS UNCORRECTABLE AT THIS TIME
;
```

Note that by convention the value of `_database_PDB_remark.id` matches the class of the REMARK record in the PDB file.

3.6.8.4. Article publication

Categories used during the publication of an article are as follows:

IUCR group

Journal housekeeping and reference entries (§3.6.8.4.1)

JOURNAL

JOURNAL_INDEX

Contents of a publication (§3.6.8.4.2)

PUBL

PUBL_AUTHOR

PUBL_BODY

PUBL_MANUSCRIPT_INCL

These categories cover both the metadata for the article (information about the article) and the text of the article itself.

3.6.8.4.1. *Journal housekeeping and citation entries*

Data items in these categories are as follows:

(*a*) JOURNAL

```
• _journal.entry_id
       → _entry.id
  _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
```

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 (_).

In mmCIF, the families of categories used to contain the text of an article for publication and to record information about the handling and processing of the article by a publisher are assigned to the IUCR category group. The name arose from the fact that CIF is sponsored by the International Union of Crystallography and several of the journals of the IUCr can handle articles submitted for publication in CIF format. However, these data items may be freely used by other publishers who wish to handle articles submitted in CIF format. The JOURNAL and JOURNAL_INDEX categories are used in the same way in the core CIF and mmCIF dictionaries, and Section 3.2.5.4 can be consulted for details.

3.6.8.4.2. *Contents of a publication*

Data items in these categories are as follows:

(*a*) PUBL
• **_publl.entry_id**
→ **_entry.id**
_publ.contact_author
_publ.contact_author_address
_publ.contact_author_email
_publ.contact_author_fax
_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.entry_id**
→ **_entry.id**
_publ_manuscript_incl.extra_defn
_publ_manuscript_incl.extra_info
_publ_manuscript_incl.extra_item

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 categories PUBL, PUBL_AUTHOR, PUBL_BODY and PUBL_MANUSCRIPT_INCL are also members of the IUCR group in the mmCIF dictionary. They are used in the same way in the core CIF and mmCIF dictionaries, and Section 3.2.5.5 can be consulted for details.

3.6.9. File metadata

As in the core CIF dictionary, information about the source and the revision history of an mmCIF may be given in the AUDIT group of categories: AUDIT, AUDIT_AUTHOR, AUDIT_CONTACT_AUTHOR and AUDIT_CONFORM (Section 3.6.9.1). However, the mmCIF dictionary differs from the core CIF dictionary in the way it expresses relationships between data blocks: instead of the core AUDIT_LINK category, mmCIF has two categories, ENTRY and ENTRY_LINK, that essentially fulfil the same role but are classified in a distinct category group (Section 3.6.9.2).

3.6.9.1. History of a data block

The categories describing the history of a data block are as follows:

AUDIT group
AUDIT
AUDIT_AUTHOR
AUDIT_CONFORM
AUDIT_CONTACT_AUTHOR

Data items in these categories are as follows:

(*a*) AUDIT
• **_audit.revision_id**
_audit.creation_date
_audit.creation_method
_audit.update_record

(*b*) AUDIT_AUTHOR
• ***_audit_author.name***
_audit_author.address

(*c*) AUDIT_CONFORM
• ***_audit_conform.dict_name***
• ***_audit_conform.dict_version***
_audit_conform.dict_location

(*d*) AUDIT_CONTACT_AUTHOR
• ***_audit_contact_author.name***
_audit_contact_author.address
_audit_contact_author.email
_audit_contact_author.fax
_audit_contact_author.phone

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

The data items in these categories are used in the same way in the mmCIF dictionary as in the core CIF dictionary (see Section 3.2.6). The data item **_audit.revision_id** has been added to the AUDIT category to provide the formal category key required by the DDL2 data model. The core data item **_audit_block_code** has been replaced by **_entry.id** (see Section 3.6.9.2).

3.6.9.2. Links between data blocks

The categories describing links between data blocks are as follows:

ENTRY group
ENTRY
ENTRY_LINK
AUDIT group
AUDIT_LINK

Data items in these categories are as follows:

(*a*) ENTRY
• **_entry.id**

(*b*) ENTRY_LINK
• **_entry_link.entry_id**
→ **_entry.id**
• **_entry_link.id**
_entry_link.details