

3. CIF DATA DEFINITION AND CLASSIFICATION

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. CLASSIFICATION AND USE OF CORE DATA

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_codен_ASTM
  _journal_codен_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_recд_copyright
  _journal_date_recд_electronic
  _journal_date_recд_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-(benzyloxy)-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*.