

3. CIF DATA DEFINITION AND CLASSIFICATION

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*.

3.2. CLASSIFICATION AND USE OF CORE DATA

Because the scope of any data value is restricted to the data block in which it resides, each data block should contain its own set of `_audit_*` data items (a requirement that is often overlooked in the construction of a CIF with multiple data blocks). The data items in the `AUDIT_LINK` category may be used to record relationships between different data blocks within the same file.

Data items in these categories are as follows:

- (a) `AUDIT`
 - `_audit_block_code`
 - `_audit_creation_date`
 - `_audit_creation_method`
 - `_audit_update_record`
- (b) `AUDIT_AUTHOR`
 - `_audit_author_address`
 - `_audit_author_name`
- (c) `AUDIT_CONFORM`
 - `_audit_conform_dict_location`
 - `_audit_conform_dict_name`
 - `_audit_conform_dict_version`
- (d) `AUDIT_CONTACT_AUTHOR`
 - `_audit_contact_author_address`
 - `_audit_contact_author_email`
 - `_audit_contact_author_fax`
 - `_audit_contact_author_name`
 - `_audit_contact_author_phone`
- (e) `AUDIT_LINK`
 - `_audit_link_block_code`
 - `_audit_link_block_description`

The `AUDIT` category provides a small set of data names suitable for identifying a data block and recording its creation date and subsequent modifications. Each data block in a CIF is introduced by a string of the form `data_xxxx`, where the block code `xxxx` is an arbitrary string. CIF offers no guidelines for choosing a block code, and there are many cases where the same string has been chosen to label data blocks in different files. The `_audit_block_code` data item is meant to encourage authors to provide a unique label for a data block. Also, as a separate data item, `_audit_block_code` has the advantage that it can be interrogated using standard CIF query tools; this is not true of the block code.

The core dictionary does not specify a procedure for choosing a unique identifier for the data block, but other dictionaries do. The modulated structures dictionary recommends specific naming procedures (Section 3.4.4.4) and the power dictionary supplies alternative data items designed to generate globally unique identifiers (Section 3.3.7.1).

Some applications modify the block code in the `data_xxxx` string. The value of `_audit_block_code` may not be changed arbitrarily to suit the convenience of external applications.

In Example 3.2.6.1, the `_audit_block_code` assigned is different from the data-block code; the creation date is expressed in the CIF date format convention of `yyyy-mm-dd` and the revision record is generated by adding material to the `_audit_update_record` field. Each addition has been prefixed with the date and initialled by the person who made the change. It is good practice to maintain a full record of any changes of substance to the contents of the data block.

Data items in the `AUDIT_AUTHOR` category record details of the author or authors of the data block. Where there is more than a single author, the names and addresses are looped. The use of these data items parallels that of the items in the `PUBL_AUTHOR` category; the difference is that the latter are used specifically to record details of authors of an article for publication. The `AUDIT_AUTHOR`

Example 3.2.6.1. *Items identifying a data block and recording its revision history.*

```
data_example

_audit_block_code           xyzzy_2002-04-05
_audit_creation_date        2002-04-05
_audit_creation_method      'SHELXL97'

_audit_update_record
; 2002-04-09 discussion added           BM
 2002-04-17 coeditor number XY1234 assigned BM
 2002-04-18 revised comment after referee report BM
;
```

Example 3.2.6.2. *The CIF dictionaries to which the data block conforms.*

```
loop_
_audit_conform_dict_name
_audit_conform_dict_version
_audit_conform_dict_location
cif_core.dic      2.3.1 .
cif_pd.dic       1.0.1 .
cif_local_my.dic 1.0
                 /usr/local/dics/my_local_dictionary
```

category refers to the creators of a CIF data block regardless of its intended purpose.

Data items in the `AUDIT_CONFORM` category describe the version of the dictionary or dictionaries that contain the definitions of the data names in the current data block. It is very helpful to provide this information, so that applications software can locate the original definitions and validate the contents of the current data block against them (Example 3.2.6.2). The dictionary identifier `_audit_conform_dict_name` is essential. The version is less important, as the dictionaries are revised in such a way as to try to retain compatibility between versions, but may occasionally be useful if changes of substance have crept in between versions. The location specified by `_audit_conform_dict_location` is useful only for local applications; in general the public register of CIF dictionaries should be used to locate dictionary files (see Section 3.1.8.3).

Data items in the `AUDIT_CONTACT_AUTHOR` category record details of the name and address of the author to be contacted concerning the contents of the data block. The use of these data items parallels that of the items in the `PUBL_CONTACT_AUTHOR` category; the difference is that the latter are used specifically to record details of the contact author of an article for publication. The `AUDIT_CONTACT_AUTHOR` category refers to the creator of a CIF data block regardless of its intended purpose.

The original purpose of a CIF, to record the data relevant to a single-crystal structure determination, was quickly extended to include the creation of an article reporting several crystal structures, as well as to powder CIFs recording information about multiple phases, modulated-structure CIFs describing superimposed structures and macromolecular CIFs recording results of multiple refinement cycles. A mechanism is required to differentiate the purpose of an individual data block and its relationship to other data blocks in the same file. This is provided by the `AUDIT_LINK` category. Example 3.2.6.3 shows how a CIF of an article for publication might show the relationships between the data blocks in the file. Note that the link references the value of `_audit_block_code` in the referenced data block, *not* the data-block header string itself (although in this example, and in Example 3.2.6.4, they have the same value).

3. CIF DATA DEFINITION AND CLASSIFICATION

Example 3.2.6.3. List of linked data blocks in a CIF.

```
data_global
_audit_block_code      global
loop_
_audit_link_block_code
_audit_link_block_description
.      'text of paper with two structures'
morA_ (1) 'structure 1 of 2'
morA_ (2) 'structure 2 of 2'
```

Example 3.2.6.4. Complementary list of linked data blocks in a secondary block.

```
data_morA_ (1)
_audit_block_code      morA_ (1)
loop_
_audit_link_block_code
_audit_link_block_description
global      'text of paper with two structures'
.      'structure 1 of 2'
morA_ (2) 'structure 2 of 2'
```

For many applications, it is enough for a statement of the links between the data blocks in a CIF to be included once only in the file, normally in the initial data block. However, for completeness and to permit consistency checking, it is best if the other data blocks in the file have complementary declarations (Example 3.2.6.4).

Current practice as described in the core dictionary restricts this reporting of links between data blocks to the contents of a single file. In principle, if `_audit_block_code` were known to have globally unique values in each distinct data block, the mechanism could be extended to permit inter-file linkage.

Appendix 3.2.1

Category structure of the core CIF dictionary

Table A3.2.1.1 provides an overview of the structure of the core CIF dictionary by informal category group and categories.

Table A3.2.1.1. Categories in the core CIF dictionary

Numbers in parentheses refer to the section of this chapter in which each category is described in detail.

ATOM group (§3.2.4.1)	DIFFRN_SCALE_GROUP (§3.2.2.2.5(d))
ATOM.SITE (§3.2.4.1.1)	DIFFRN_SOURCE (§3.2.2.2.2(d))
ATOM.SITES (§3.2.4.1.2)	DIFFRN_STANDARD_REFLN (§3.2.2.2.5(e))
ATOM.TYPE (§3.2.4.1.3)	DIFFRN_STANDARDS (§3.2.2.2.5(f))
AUDIT group (§3.2.6)	EXPTL group (§3.2.2.3)
AUDIT (§3.2.6(a))	EXPTL (§3.2.2.3(a))
AUDIT_AUTHOR (§3.2.6(b))	EXPTL_CRYSTAL (§3.2.2.3(b))
AUDIT_CONFORM (§3.2.6(c))	EXPTL_CRYSTAL_FACE (§3.2.2.3(c))
AUDIT_CONTACT_AUTHOR (§3.2.6(d))	GEOM group (§3.2.4.3)
AUDIT_LINK (§3.2.6(e))	GEOM (§3.2.4.3.1(a))
CELL group (§3.2.2.1)	GEOM_ANGLE (§3.2.4.3.1(b))
CELL (§3.2.2.1(a))	GEOM_BOND (§3.2.4.3.1(c))
CELL_MEASUREMENT_REFLN (§3.2.2.1(b))	GEOM_CONTACT (§3.2.4.3.1(d))
CHEMICAL group (§3.2.4.2)	GEOM_HBOND (§3.2.4.3.1(e))
CHEMICAL (§3.2.4.2.1(a))	GEOM_TORSION (§3.2.4.3.1(f))
CHEMICAL_CONN_ATOM (§3.2.4.2.2(a))	JOURNAL group (§3.2.5.4)
CHEMICAL_CONN_BOND (§3.2.4.2.2(b))	JOURNAL (§3.2.5.4(a))
CHEMICAL_FORMULA (§3.2.4.2.1(b))	JOURNAL_INDEX (§3.2.5.4(b))
CITATION group (§3.2.5.1)	PUBL group (§3.2.5.5)
CITATION (§3.2.5.1(a))	PUBL (§3.2.5.5(a))
CITATION_AUTHOR (§3.2.5.1(b))	PUBL_AUTHOR (§3.2.5.5(b))
CITATION_EDITOR (§3.2.5.1(c))	PUBL_BODY (§3.2.5.5(c))
COMPUTING group (§3.2.5.2)	PUBL_MANUSCRIPT_INCL (§3.2.5.5(d))
COMPUTING (§3.2.5.2)	REFINE group (§3.2.3.1)
DATABASE group (§3.2.5.3)	REFINE (§3.2.3.1(a))
DATABASE (§3.2.5.3)	REFINE_LS_CLASS (§3.2.3.1(b))
DIFFRN group (§3.2.2.2)	REFLN group (§3.2.3.2)
DIFFRN (§3.2.2.2.1)	REFLN (§3.2.3.2.1)
DIFFRN_ATTENUATOR (§3.2.2.2.2(a))	REFLNS (§3.2.3.2.2(a))
DIFFRN_DETECTOR (§3.2.2.2.4)	REFLNS_CLASS (§3.2.3.2.2(b))
DIFFRN_MEASUREMENT (§3.2.2.2.3(a))	REFLNS_SCALE (§3.2.3.2.2(c))
DIFFRN_ORIENT_MATRIX (§3.2.2.2.3(b))	REFLNS_SHELL (§3.2.3.2.2(d))
DIFFRN_ORIENT_REFLN (§3.2.2.2.3(c))	SYMMETRY group (§3.2.4.4)
DIFFRN_RADIATION (§3.2.2.2.2(b))	SPACE_GROUP (§3.2.4.4.2(a))
DIFFRN_RADIATION_WAVELENGTH (§3.2.2.2.2(c))	SPACE_GROUP_SYMOP (§3.2.4.4.2(b))
DIFFRN_REFLN (§3.2.2.2.5(a))	SYMMETRY (§3.2.4.4.1(a))
DIFFRN_REFLNS (§3.2.2.2.5(b))	SYMMETRY_EQUIV (§3.2.4.4.1(b))
DIFFRN_REFLNS_CLASS (§3.2.2.2.5(c))	VALENCE group (§3.2.4.5)
	VALENCE_PARAM (§3.2.4.5(a))
	VALENCE_REF (§3.2.4.5(b))

References

- Busing, W. R. & Levy, H. A. (1967). *Angle calculations for 3- and 4-circle X-ray and neutron diffractometers*. *Acta Cryst.* **22**, 457–464.
- Hall, S. R. (1981). *Space-group notation with an explicit origin*. *Acta Cryst.* **A37**, 517–525; erratum (1981), **A37**, 921.
- Hall, S. R. & Grosse-Kunstleve, R. W. (2001). *International tables for crystallography*, Vol. B, *Reciprocal space*, edited by U. Shmueli, 2nd ed., Appendix A1.4.2.3. Dordrecht: Kluwer Academic Publishers.
- Klyne, W. & Prelog, V. (1960). *Description of steric relationships across single bonds*. *Experientia*, **16**, 521–523.