

3. CIF DATA DEFINITION AND CLASSIFICATION

Table 3.1.10.1. High-level grouping of categories by dictionary

Category groups are organized into families by common function and purpose.

cif_core.dic	cif_pd.dic	cif_ms.dic	cif_rho.dic	mmcif_std.dic	cif_img.dic	cif_sym.dic
<i>(a)</i> Experimental measurements						
CELL DIFFRN EXPTL	PD_CALIB PD_CHAR PD_DATA PD_INSTR PD_MEAS PD_PREP PD_SPEC	CELL DIFFRN EXPTL		CELL DIFFRN EXPTL	ARRAY AXIS DIFFRN	
<i>(b)</i> Analysis						
REFINE REFLN	PD_CALC PD_PEAK PD_PROC REFLN	REFINE REFLN		PHASING REFINE REFLN		
<i>(c)</i> Structure						
ATOM CHEMICAL		ATOM	ATOM	ATOM CHEMICAL CHEM_COMP CHEM_LINK ENTITY GEOM		
GEOM	PD_PHASE	GEOM		STRUCT SYMMETRY VALENCE		SPACE_GROUP
SYMMETRY VALENCE		SYMMETRY				
<i>(d)</i> Publication						
CITATION COMPUTING DATABASE JOURNAL PUBL				CITATION COMPUTING DATABASE JOURNAL PUBL SOFTWARE		
<i>(e)</i> File metadata						
AUDIT	PD_BLOCK	AUDIT		AUDIT		

and practices of each field. Here we provide a general framework within which the category groups of each separate dictionary may be described.

3.1.10.1. Categories and category groups

The only formal unit of classification common to all CIF dictionaries is the *category*. For example, in the core CIF dictionary information about the chemical and physical properties of the different atomic species in a crystal cell is collected in a few data names such as `_atom_type_oxidation_number` which belong to the same category, in this case the `ATOM_TYPE` category. As described in Section 3.1.5.3, it is conventional (although not mandatory) that CIF data names begin with components corresponding to the name of the category to which they belong.

The term *category* as used in CIF dictionaries has a technical meaning which constrains its normal use in grouping items that are understood to have a 'natural' relationship. In a CIF, only items belonging to the same category may appear together in the same looped list. This means, for example, that data items describing collective properties of the atom sites in the lattice (such as the number of atoms of each atomic species in the unit cell) must be assigned to a different category from the data items that describe the properties of the individual sites. Hence the properties of individual sites (such as the positional coordinates defined by `_atom_site_fract_x` etc.) belong to the `ATOM_SITE` category, while the transformation matrix between Cartesian and fractional components (expressed by a collection of data names such as

`_atom_sites_fract_tran_matrix_11`) belong to the `ATOM_SITES` category. Clearly, the category names have been chosen to be similar to reflect their close relationship, while the `EXPTL` category containing data names such as `_exptl_crystal_colour` is named quite differently. It is natural to wish to describe related categories in a common higher level of classification, and indeed *category groups* exist as formal components of DDL2-structured dictionaries. We shall, however, refer informally to 'category groups' in discussions of DDL1 dictionaries as collections of categories with a close relationship that is usually implicit in their names.

3.1.10.2. Overview of category classification

Table 3.1.10.1 provides an informal classification at a high level of the category groups represented in each of the CIF dictionaries in this volume. Related category groups are clustered within the table in families sharing a common function. The five families (*a*) to (*e*) in Table 3.1.10.1 refer to: the crystallographic experiment itself; the processing and analysis of data from the experiment; the derived structural model; the reporting and publication of the results; and general auditing of the file itself, its purpose, authorship, history and links to other data sets, *i.e.* the file metadata. Detailed discussions of the individual categories (and formal category groups for DDL2 dictionaries) will be found in the relevant chapters in the rest of this part of the volume.

Table 3.1.10.1 shows the different characters of the seven dictionaries. The macromolecular dictionary (`mmcif_std.dic`; Chapter 4.5) contains an embedded version of the core dictionary