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
(*a*) Experimental measurements						
					ARRAY	
					AXIS	
CELL		CELL		CELL		
DIFFRN		DIFFRN		DIFFRN	DIFFRN	
EXPTL		EXPTL		EXPTL		
	PD_CALIB					
	PD_CHAR					
	PD_DATA					
	PD_INSTR					
	PD_MEAS					
	PD_PREP					
	PD_SPEC					
(*b*) Analysis						
	PD_CALC					
	PD_PEAK					
	PD_PROC					
				PHASING		
REFINE		REFINE		REFINE		
REFLN	REFLN	REFLN		REFLN		
(*c*) Structure						
ATOM		ATOM	ATOM	ATOM		
CHEMICAL				CHEMICAL		
				CHEM_COMP		
				CHEM_LINK		
				ENTITY		
GEOM		GEOM		GEOM		
	PD_PHASE					
				STRUCT		
SYMMETRY		SYMMETRY		SYMMETRY		SPACE_GROUP
VALENCE				VALENCE		
(*d*) Publication						
CITATION				CITATION		
COMPUTING				COMPUTING		
DATABASE				DATABASE		
JOURNAL				JOURNAL		
PUBL				PUBL		
				SOFTWARE		
(*e*) File metadata						
AUDIT		AUDIT		AUDIT		
	PD_BLOCK					

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

(cif_core.dic; Chapter 4.1) in DDL2 format and so includes all the categories defined in the core. However, it extends the description of the structural model extensively by introducing families of categories for the description of chemical components of a macromolecular structure (ENTITY) and for the detailed description of the structure itself (STRUCT). New categories are also introduced to describe the phasing of the structure and the SOFTWARE category allows the inclusion of more details of computational techniques than the core COMPUTING category does.

The other dictionaries are purely extensions which either introduce new data names (and occasionally new categories) into existing category groups or, where necessary, introduce completely new groups of categories.

The powder dictionary (cif_pd.dic; Chapter 4.2) contains several new category groups reflecting the need for substantially different methods of describing the experiment and analysing the data, as well as a need for the structural model to be able to handle multiple crystalline phases. The modulated structures dictionary (cif_ms.dic; Chapter 4.3) introduces no new category groups, but does introduce several new data names and categories within the existing framework. The electron density dictionary (cif_rho.dic; Chapter 4.4) introduces two new categories within an existing category group. The image CIF dictionary (cif_img.dic; Chapter 4.6) has several new categories that characterize arrays of data from two-dimensional X-ray detectors and the consequent detailed descriptions of the relevant axes within the experimental setup. The symmetry dictionary (cif_sym.dic; Chapter 4.7) was commissioned specifically to replace the symmetry categories in the core dictionary with a more detailed treatment.

References

Bray, T., Paoli, J. & Sperberg-McQueen, C. M. (1998). *Extensible Markup Language (XML) 1.0*. W3C Recommendation 10-February-1998. http://www.w3.org/TR/1998/REC-xml-19980210.

CrossRef (2004). *Query spec.* http://www.crossref.org/03libraries/25query_spec.html.

Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *The Crystallographic Information File (CIF): a new standard archive file for crystallography. Acta Cryst.* A**47**, 655–685.

Hall, S. R., Spadaccini, N., Castleden, I. R., du Boulay, D. & Westbrook, J. D. (2002). *StarDDL: towards the unification of Star dictionaries. Acta Cryst.* A**58** (Suppl.), C256.

W3C (2004). *Extensible Markup Language (XML)*. http://www.w3c.org/XML/.