Table 3.4.2.1. *Category groups defined in the msCIF dictionary*

The groups are listed in the order in which they are described in this chapter.

Section	Category group	Subject covered
(*a*) Experimental measurements		
3.4.3.1.1	CELL	The unit cell, especially cell subsystems for composite structures
3.4.3.1.2	DIFFRN	Multi-dimensional diffraction pattern
3.4.3.1.3	EXPTL	Details of the experiment specific to modulated and composite structures
(*b*) Analysis		
3.4.3.2.1	REFINE	Refinement procedures
3.4.3.2.2	REFLN	Reflection measurements indexed in higher-dimensional space
(*c*) Atomicity, chemistry and structure		
3.4.3.3.1 to 3.4.3.3.3	ATOM	Atom sites in a modulated structure
3.4.3.3.4	GEOM	Geometry of a structure in superspace
3.4.3.3.5	SYMMETRY	Symmetry information
(*d*) File metadata		
3.4.3.4	AUDIT	The structure of the CIF

(Chapter 4.1). This means that the dictionary defines items that are basically related to single-crystal data. The close relationship between the msCIF and core dictionaries has led to synergies between and benefits for both dictionaries.

The design of the msCIF dictionary had two objectives: (i) it should be as functional as possible, *i.e.* as little information in an msCIF as possible should be given as unstructured text; (ii) it should be possible to include even the oldest modulated and composite structures in an msCIF, even if the way they were reported did not follow the guidelines used now.

There were two major difficulties in implementing the msCIF dictionary. Firstly, the number of additional wave vectors used to index a diffraction diagram is theoretically not limited. Secondly, a CIF containing information about a modulated or composite structure should, in general, be composed of several (related) data blocks. As CIF definitions do not at present include vectors or matrices as distinct types of data values, an arbitrary upper limit of 11 was assigned to the dimension of superspace to limit the number of new data names. Linking between data blocks is handled by using recommended values for items in the AUDIT and AUDIT_LINK categories, like those used in the powder CIF (pdCIF) dictionary (Chapters 3.3 and 4.2).

An additional problem arises when special (ideal) modulation functions are considered. Although periodic modulations are normally parameterized by Fourier series, in certain cases it is convenient to use discontinuous functions which lead to a severe reduction in the number of structural parameters. The shape of these functions is not restricted and new materials could require new functions. Given that it is not possible at this moment to define logical or mathematical relations between data values within a CIF [although an initiative for including algorithms in the definitions of CIF dictionaries has been proposed by Spadaccini *et al.* (2000)], general functions cannot be defined and therefore the type of special functions included in the msCIF dictionary are those implemented in the most widely used program, *JANA2000* (Petříček & Dušek, 2000). They only apply to one-dimensional modulations and are sawtooth displacive functions and occupational crenel functions. Both functions define discontinuous occupational atomic domains and are normally combined with (smoother) atomic modulation functions (involving atom positions and/or thermal parameters) that are expressed by Fourier series. Because of the discreteness of the atomic domains, members of the set of harmonic functions used to expand these series are no longer mutually orthogonal as they are only defined within each atomic domain and not in the (internal space) interval $[0, 1]$. As a consequence, severe correlation effects among the coefficients of the Fourier series are expected. A solution for this problem lies in the selection and orthogonalization of a set of basic functions (Petříček *et al.*, 1995). The atomic modulation functions are then expressed as linear combinations of an orthogonal basis whose elements are specific combinations of harmonic functions. Discontinuous atomic domains are being increasingly used in composite materials, in some cases revealing that considering such materials as composites or modulated structures is a matter of convenience (Elcoro *et al.*, 2003; Pérez-Mato *et al.*, 2003).

CIFs that conform to the msCIF dictionary are highly itemized for the human reader, but have a strong relational structure even though the dictionary itself is written in DDL1.

The major drawbacks of the dictionary are:

(i) Items describing the superspace symmetry should be reconsidered and perhaps included within the symmetry CIF (symCIF) dictionary (Chapters 3.8 and 4.7).

(ii) There is still some information in an msCIF that cannot be interpreted by a computer [*e.g.* rigid rotations and translations around (along) noncrystallographic axes cannot be parsed, since the description of the axes is textual].

(iii) A full description of the modulation in terms of orthogonalized functions (used when the atomic domains are discrete) is not supported yet.

3.4.3. Arrangement of the dictionary

The msCIF dictionary detailed in Chapter 4.3 includes 19 new categories. Another 18 already exist in the core CIF dictionary, but include new items (16) or revised definitions (2). The category structure of the msCIF dictionary is summarized in Table 3.4.2.1 and is listed in full in Appendix 3.4.1. The appendix also lists for each category the section of this chapter in which the category is described.

Many of the modifications to categories that already exist in the core CIF dictionary result from the need to use more than three integer indices to label the diffracted intensities (in the cases of CELL, DIFFRN_REFLN, DIFFRN_REFLNS, DIFFRN_STANDARD_REFLN and EXPTL_CRYSTAL_FACE) or the need to use superspace symmetry (in the cases of GEOM_ANGLE, GEOM_BOND, GEOM_CONTACT, GEOM_TORSION, SPACE_GROUP and SPACE_GROUP_SYMOP). Apart from the categories that describe the atomic modulation functions, there are two that are specific to composite structures (CELL_SUBSYSTEM and CELL_SUBSYSTEMS).

The rest of this section summarizes the contents of the dictionary, organized by categories within the functional groups outlined in Table 3.1.10.1. As in the other chapters in this part of the volume, the classification is under the headings *Experimental measurements* (Section 3.4.3.1), *Analysis* (Section 3.4.3.2), *Atomicity, chemistry and structure* (Section 3.4.3.3) and *File metadata* (Section 3.4.3.4). The msCIF dictionary adds no new data items concerned with the publication or reporting of structures to those already present in the core CIF dictionary.

The data items within each category are listed in the detailed commentary below. Where relevant, data items that represent a unique identifier for a looped list ('category keys') are listed first and are marked by a bullet (•). The remaining data items in each category are listed alphabetically.