

3.2. Classification and use of core data

BY S. R. HALL, P. M. D. FITZGERALD AND B. MCMAHON

3.2.1. Introduction

This chapter is concerned with the classification and organization of data items defined in the core CIF dictionary (Chapter 4.1). The core dictionary, as its name suggests, is central to the definition of data items found in most CIFs. It defines the measured and derived items common to most crystallographic experiments, analyses and publications, and, in particular, those items characterizing a classical single-crystal X-ray diffraction determination of a small-molecule or inorganic structure. As the nature of crystallographic studies evolves, so do the data items needed to describe them. New data names are introduced as needed to describe new techniques or technologies or simply to provide more details of subjects already covered. In addition, the developers of specialist dictionaries may find that some of the items they define have a wider application and propose that these items be added to the core dictionary instead.

Core data items are defined with two formalisms. The core dictionary, as presented in Chapter 4.1, defines core data items exclusively using the data definition language DDL1 (described in Chapter 2.5). However, core data items are also embedded within the macromolecular CIF dictionary presented in Chapter 4.5 using the data definition language DDL2 (described in Chapter 2.6). Because the revision cycles of the core and mmCIF dictionaries are not synchronized, at any one time the mmCIF dictionary may not include the complete set of data items in the current core dictionary. The mmCIF dictionary described in this volume includes the full content of core CIF dictionary version 2.3.1, also described in this volume.

The discussion in this chapter will concentrate on the current DDL1 version of the core dictionary (version 2.3, released on 4 October 2003 and reissued with minor amendments as version 2.3.1 in this volume). There will be some discussion of the more formal approach to the classification of data items that DDL2 permits.

In accordance with the scheme given in Table 3.1.10.1, groups of categories of data items in the core dictionary will be classified under the headings *Experimental measurements* (Section 3.2.2), *Analysis* (Section 3.2.3), *Atomicity, chemistry and structure* (Section 3.2.4), *Publication* (Section 3.2.5) and *File metadata* (Section 3.2.6). To help the reader relate the thematic order of the discussion of these categories to the alphabetic layout of the dictionary, the category structure of the core dictionary is summarized in Table 3.2.1.1 and is listed in full in Appendix 3.2.1. The appendix also lists for each category the section of this chapter in which the category is described.

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

Table 3.2.1.1. *Category groups defined in the core CIF dictionary*

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

Section	Category group	Subject covered
<i>(a) Experimental measurements</i>		
3.2.2.1	CELL	Unit cell
3.2.2.2	DIFFRN	Diffraction experiment
3.2.2.3	EXPTL	Experimental conditions
<i>(b) Analysis</i>		
3.2.3.1	REFINE	Refinement procedures
3.2.3.2	REFLN	Reflection measurements
<i>(c) Atomicity, chemistry and structure</i>		
3.2.4.1	ATOM	Atom sites
3.2.4.2	CHEMICAL	Chemical properties and nomenclature
3.2.4.3	GEOM	Geometry of atom sites
3.2.4.4	SYMMETRY	Symmetry information
3.2.4.5	VALENCE	Bond-valence information
<i>(d) Publication</i>		
3.2.5.1	CITATION	Bibliographic references
3.2.5.2	COMPUTING	Computational details of the experiment
3.2.5.3	DATABASE	Database information
3.2.5.4	JOURNAL	Journal housekeeping
3.2.5.5	PUBL	Contents of a published article
<i>(e) File metadata</i>		
3.2.6	AUDIT	Dictionary maintenance and identification

category keys are defined more formally in the mmCIF dictionary (see Chapter 2.6 and the discussion of categories in Section 3.1.6.4). The remaining data items in each category are listed alphabetically.

3.2.2. Experimental measurements

Crystallographic archive files predating CIF were often constructed to serve the purposes of a particular software program or suite and stored the data generated by an experiment without providing a full record of the conditions under which the data were obtained. This is not unique to crystallography: many data formats make no provision for the metadata – information about the procedures for gathering and analysing data – that give context and in many cases significance to the numeric values. A specific goal of the design of CIF was to treat such supporting information as essential elements of the whole collection of information relating to a structure determination, rather than as optional and poorly defined metadata. There are therefore many categories in the core dictionary that relate to experimental conditions and apparatus, and these categories are discussed in this section. They include the categories in the DIFFRN group describing the traditional crystallographic diffraction experiment (typically a single-crystal laboratory-based X-ray determination, but increasingly including synchrotron experiments and experiments using other radiation types). There are also categories that describe and characterize the crystal used in the experiment and those that characterize the unit cell, since the experimental determination of the cell parameters is an essential part of the full structure-determination experiment.

Affiliations: SYDNEY R. HALL, School of Biomedical and Chemical Sciences, University of Western Australia, Crawley, 6009, Australia; PAULA M. D. FITZGERALD, Merck Research Laboratories, Rahway, New Jersey, USA; BRIAN MCMAHON, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.