2.6. Specification of a relational dictionary definition language (DDL2)

BY J. D. WESTBROOK, H. M. BERMAN AND S. R. HALL

2.6.1. Introduction

The dictionary definition language version 2 (DDL2) presented here extends the DDL1 version (Hall & Cook, 1995) currently used by the IUCr for the description of data items common to all crystallographic studies (*i.e.* core items). The DDL2 extensions were introduced primarily to address two issues arising during the development of a CIF dictionary for the terminology of macromolecular crystallography: the need to accurately describe the hierarchical nature of macromolecular structure and associated structural features, and the desire to encode dictionary definitions in a manner that would permit more detailed software-driven validation.

The decision not to use DDL1 for the macromolecular CIF dictionary (mmCIF) was not made lightly. The Working Group responsible for the development of the mmCIF dictionary spent three years building a dictionary version within a DDL1 framework. When presented at the first mmCIF workshop in York in 1993 (Chapter 1.1), it was criticized as lacking the rigour to be usefully interpreted by software. In particular, the draft dictionary lacked machine-interpretable relationships between the components of macromolecular structure, the components of structure and structural properties, and the components of structure and the experimental description. These relationships are important to a complete description of macromolecular structural data, and need to be present in a dictionary in a form that permits software to navigate and validate them.

Following the York workshop, the Working Group set about redesigning the framework of the data model used to organize the dictionary definitions. Initially there was significant interest in adopting a more object-oriented data model so as to match closely the object-oriented characteristics of macromolecular structure. However, an object-oriented model would depart significantly from the organization of the core CIF dictionary based on DDL1, and interoperability between the two approaches would be likely to be problematic.

The new DDL concepts were presented at the Brussels mmCIF workshop in 1994. This approach, which later became known as DDL2 (Westbrook & Hall, 1995), employs a largely relational data model that adhered more closely to the data model implicit in the core DDL1. DDL2 added new elements to expand the DDL1 concept of a data category. Categories in DDL2 are fully realized definitional elements that have their own set of attributes (*e.g.* definitions and examples). DDL2 also added data elements to explicitly define parent–child relationships between data items within a hierarchy of categories. It was demonstrated that using this conservative relational data model it was possible to accurately describe the content in the mmCIF dictionary.

Affiliations: JOHN D. WESTBROOK and HELEN M. BERMAN, Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers, The State University of New Jersey, Department of Chemistry and Chemical Biology, 610 Taylor Road, Piscataway, NJ 08854-8087, USA; SYDNEY R. HALL, School of Biomedical and Chemical Sciences, University of Western Australia, Crawley, Perth, WA 6009, Australia.

Example 2.6.2.1. *Typical mmCIF definition save frame.*

```
save__exptl.details
  _item_description.description
; Any special information about the experimental work
  prior to the intensity measurement. See also
  _exptl_crystal.preparation.
;
  _item.name                  '_exptl.details'
  _item.category_id           exptl
  _item.mandatory_code        no
  _item_aliases.alias_name    '_exptl_special_details'
  _item_aliases.dictionary    cif_core.dic
  _item_aliases.version       2.0.1
  _item_type.code             text
save_
```

In view of the importance of maintaining continuity between small- and large-molecule crystallography, the extensions in DDL2 have been introduced in a manner that provides the greatest degree of backward compatibility with applications and dictionaries developed on the core DDL1. Like DDL1, DDL2 dictionaries and data files are fully compliant with the underlying syntax rules of the Self-defining Text Archive and Retrieval (STAR) File (Hall, 1991; Hall & Spadaccini, 1994). Although DDL2 uses a different convention to name data items, an alias feature in DDL2 is used to maintain a correspondence with the published data-item names in the core CIF dictionary (Hall *et al.*, 1991; Chapter 4.1).

Since its introduction in 1994, DDL2 has been used to develop the mmCIF dictionary and this has been adopted as the data exchange standard of the Protein Data Bank, the international repository of three-dimensional macromolecular structure data (see Chapter 5.5).

2.6.2. The DDL2 presentation

The syntax used in mmCIF data files and DDL2 dictionaries is a subset of the STAR syntax. Definitions in DDL2 dictionaries are encapsulated in named save frames (see Chapter 2.1 for a description of the STAR File syntax). A save frame is a syntactical element that begins with the `save_` directive and is terminated by another `save_` directive. Save frames are named by appending a text string to the `save_` directive. In the mmCIF dictionary, save frames are used to encapsulate item and category definitions. The mmCIF dictionary is composed of a data block containing thousands of save frames, where each save frame contains a different definition. Save frames appear in DDL2 dictionaries but they are not used in data files. Save frames may not be nested.

DDL2 dictionary definitions typically contain a small number of items that specify the essential features of the item. Example 2.6.2.1 shows a save frame containing the mmCIF definition of the data item `_exptl.details`.

The example definition includes: a description or text definition, the name and category of the item, a code indicating that the item is optional (not mandatory), the name of a related definition in the core CIF dictionary, and a code specifying that the data type is text. A more detailed description of the elements of the dictionary definitions is presented in the following sections.

references