

2. CONCEPTS AND SPECIFICATIONS

2.6.3. Overview of the elements of DDL2

The elements of DDL2 provide the organizational framework for building data dictionaries like mmCIF. The role of the DDL is to define which data items may be used to construct the definitions in the data dictionary, and also to define the relationships between these defining data items. The DDL2 attributes are defined in the dictionary presented in Chapter 4.10.

A dictionary language contains no specific information about a discipline, such as macromolecular crystallography; rather, it defines the data items that can be used to describe a discipline. The contents of the mmCIF dictionary are metadata, or data about data. The contents of the DDL are meta-metadata, the data defining the metadata. By design DDL2, like its predecessor DDL1, is quite generic. It defines data items that describe the general features of a data item like a textual description, a data type, a set of examples, a range of permissible values, or perhaps a discrete set of permitted values. Consequently, data modelling using a DDL can be applied in many fields.

The lowest level of organization provided by DDL2 is the description of an individual data item. Collections of related data items are organized in categories. Categories are essentially tables in which each repetition of the group of related items adds a row. The terms category and data item are used here in order to conform with the previous use of these terms by STAR and CIF applications; these terms could be replaced by relation and attribute (or table and column) commonly used to describe the relational model which underlies DDL2.

Within a category, the set of data items determining the uniqueness of their group are designated as key items in the category.

No data-item group in a category is allowed to have a set of duplicate values of its key items. Each data item is assigned membership in one or more categories. Parent-child relationships may be specified for items belonging to multiple categories. A parent-child relationship identifies cases in which the same data item, often an important identifier, occurs in different categories. These relationships permit the expression of the very complicated hierarchical data structures required to describe macromolecular structure.

Other levels of organization in addition to category are also supported. Related categories may be collected together in category groups and parent relationships may be specified for these groups. This higher level of association provides a vehicle for organizing a large complicated collection of categories into smaller more relevant and potentially interrelated groups. This effectively provides a chaptering mechanism for large and complicated dictionaries, like mmCIF. Within the level of a category, subcategories of data items may be defined among groups of related data items. The subcategory provides a mechanism for identifying, for example, that the data items month, day, and year collectively define a date.

For categories, subcategories and items, methods may be specified. Methods are computational procedures that are defined and expressed in a programming language (*e.g.* C/C++, Perl or Java) and stored within a dictionary. Among other things, these dictionary methods may be used to calculate a missing value or to check the validity of a particular value.

The highest levels of data organization provided by DDL2 are the data block and the dictionary. The dictionary level collects a set of related definitions into a single unit and provides the attributes for a detailed revision history of the collection.

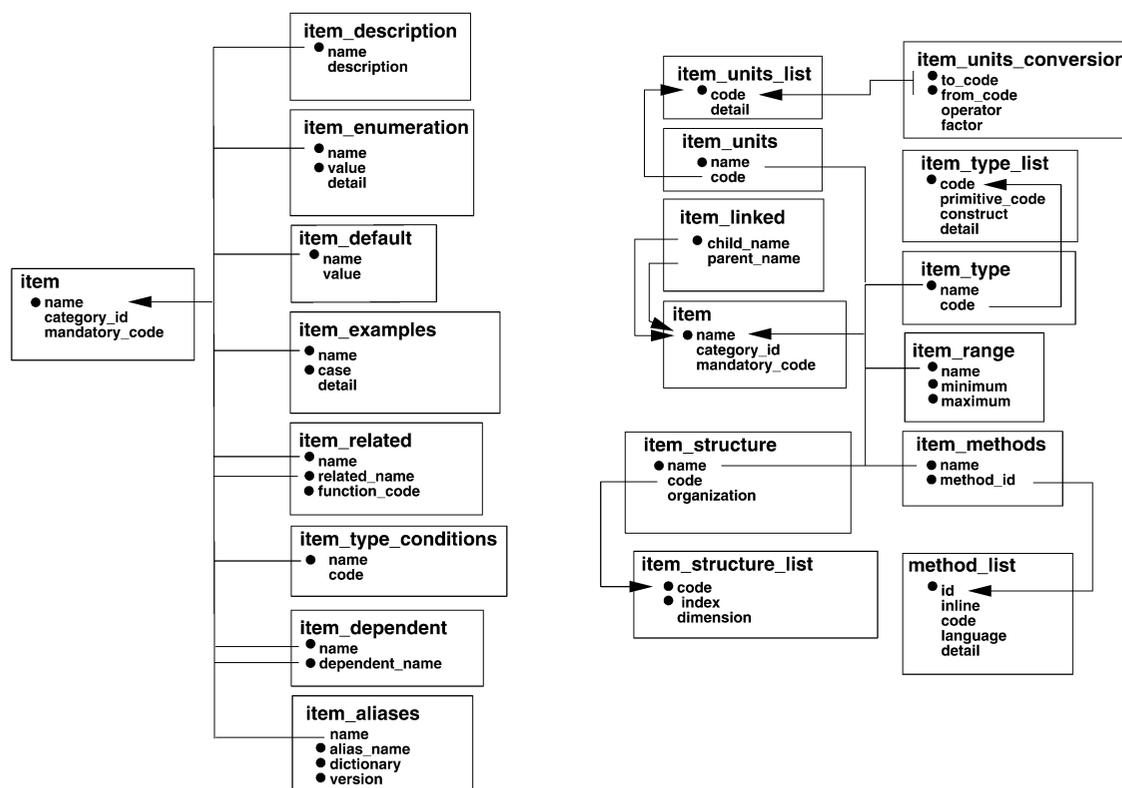


Fig. 2.6.4.1. DDL2 attributes used to specify item information. In this figure, boxes surround data items belonging to the same category. Category identifiers are given in a large typesize and item names are given in a smaller typesize. Parent-child relationships are specified by lines connecting data items with the arrow pointing at the parent item. Key items within a category are marked with a bullet.