(http://www.iucr.org/iucr-top/cif) should be consulted for current practice. However, a short overview of the existing procedures is helpful in describing how the community can participate in extending the standard.

3.1.3.1. Dictionary maintenance groups

Each published dictionary authorized by COMCIFS has a group of specialists appointed or invited to extend and maintain the dictionary to serve the changing needs of the subdiscipline that sponsors the dictionary. Members of these dictionary maintenance groups (DMGs) may suggest extensions or corrigenda on their own initiative or may pass on requests for extensions from individual crystallographers. A DMG will typically debate and review any suggested amendments and produce a draft revised dictionary for approval by COMCIFS.

3.1.3.2. mmCIF review cycle

The macromolecular CIF dictionary covers a very broad and active field, and a more formal procedure exists for the submission and review of proposed extensions. Possible new definitions are submitted using *pro forma* dictionary templates to a member of an editorial board appointed by the mmCIF dictionary maintenance group. Accepted proposals are approved by the DMG and released for general community review in provisional extension dictionaries as circumstances require. The extension dictionary is revised as necessary and is finally incorporated within the parent mmCIF dictionary after COMCIFS approval has been granted.

3.1.3.3. New dictionaries

A completely new dictionary to cover a subdiscipline not otherwise catered for may be commissioned by COMCIFS or may arise from community action, occasionally sponsored by an IUCr Commission. A working group is appointed to create the dictionary and relevant example files or software. The working group is expected to test the new dictionary extensively within its own community before submitting it to COMCIFS for initial approval. It is the responsibility of COMCIFS to check the dictionary for technical consistency and for compatibility with related dictionaries. COMCIFS may refer the dictionary back to the working group for further revisions. When the dictionary finally receives formal COMCIFS approval and is published, a dictionary maintenance group is formed to promote its further development (Section 3.1.3.1). The DMG usually includes one or more members of the initial working group and at least one voting member of COMCIFS.

3.1.4. Choice of data model

The following sections of this chapter describe the technical considerations in defining data items within a dictionary. Fundamental to this is the *data model* on which the dictionary is based. The STAR File upon which CIF is based is a very versatile data format and can accommodate a variety of data models. However, the use within CIF of a single level of looping enforces a rather flat data structure and a typical CIF maps most easily onto a relational database model. This is implicit in DDL1, which assigns different attributes to data items depending on whether they appear in data loops or not. Generally speaking, one may consider a list header and its associated data values as the head and body of a table of data values. The list header (or equivalently the table head) identifies the data items ranged by column within the table. For the dictionary entries relating to the data names in the list header, the `_category` attribute collects together data items which may be looped together in the same table, and the `_list_reference`, `_list_mandatory` and `_list_uniqueness` attributes work together to indicate the data items that must be present and collectively have a unique value to identify a specific row in a table of values.

Example 3.1.4.1. *Core dictionary definitions for the atom-site labels and bond distances in a CIF table of molecular geometry.*

```
data_geom_bond_atom_site_label_
    loop_ _name
                        '_geom_bond_atom_site_label_1'
                        '_geom_bond_atom_site_label_2'
    _category                       geom_bond
    _type                           char
    _list                           yes
    _list_mandatory                 yes
    _list_link_parent             '_atom_site_label'
    _definition
;     The labels of two atom sites that form a bond.
      These must match labels specified as
      _atom_site_label in the atom list.
;

data_geom_bond_distance
    _name                         '_geom_bond_distance'
    _category                       geom_bond
    _type                           numb
    _type_conditions                esd
    _list                           yes
    _list_reference     '_geom_bond_atom_site_label_'
    _enumeration_range              0.0:
    _units                          A
    _units_detail                 'angstroms'
    _definition
;     The intramolecular bond distance in angstroms.
;
```

For example, the following example from the core CIF dictionary (Chapter 4.1) shows a table of bond distances. The dictionary definitions are given in Example 3.1.4.1.

```
loop_
_geom_bond_atom_site_label_1
_geom_bond_atom_site_label_2
_geom_bond_distance
  O1   C2    1.342(4)
  O1   C5    1.439(3)
  C2   C3    1.512(4)
  C2   O21   1.199(4)
  C3   N4    1.465(3)
  C3   C31   1.537(4)
  C3   H3    1.00(3)
  N4   C5    1.472(3)
```

Within the dictionary, entries for all of `_geom_bond_distance`, `_geom_bond_atom_site_label_1` and `_geom_bond_atom_site_label_2` share the same `_category` attribute, namely 'geom_bond'. (In the rest of this chapter, as elsewhere in the volume, we refer to categories by the upper-case form of their category attribute values; here, therefore, we are referring to the GEOM_BOND category.) The entry for `_geom_bond_distance` has a `_list_reference` value of `'_geom_bond_atom_site_label_'` indicating the data names that may be used to identify this particular table. The trailing underscore in this example indicates that all matching data names must be considered as components of a compound identifier; for this case the matching data names are `'_geom_bond_atom_site_label_1'` and `'_geom_bond_atom_site_label_2'`. The dictionary entry for `_geom_bond_atom_site_label_` has a `_list_mandatory` value of `yes`, indicating that these data items *must* be present within the table. In this way, the attributes specify the unique key within a database table (in this case, the key has multiple components: the labels of both contributing atom sites).

However, the mapping onto a relational database is not exact. In some cases CIFs may present data from a single category across