References do not have to be to entries in databases of sequences: any external database can be referenced. For other kinds of databases, only the data items in the STRUCT_REF category would usually be used. The element of the structure that is referenced could be either an entity or a biological unit, that is, either a building block of the structure or a structurally meaningful assembly of those building blocks. Since the identification of the part of the structure being linked to an entry in an external database can be made using either `_struct_ref.biol_id` or `_struct_ref.entity_id`, and since any part of the structure could be linked to any number of entries in external databases, the data item `_struct_ref.id` was introduced as the category key.

3.6.7.5.7. *β-sheets*

Data items in these categories are as follows:

(*a*) STRUCT_SHEET

```
• _struct_sheet.id
  _struct_sheet.details
  _struct_sheet.number_strands
  _struct_sheet.type
```

(*b*) STRUCT_SHEET_TOPOLOGY

```
• _struct_sheet_topology.range_id_1
        → _struct_sheet_range.id
• _struct_sheet_topology.range_id_2
        → _struct_sheet_range.id
• _struct_sheet_topology.sheet_id
        → _struct_sheet.id
  _struct_sheet_topology.offset
  _struct_sheet_topology.sense
```

(*c*) STRUCT_SHEET_RANGE

```
• _struct_sheet_range.id
• _struct_sheet_range.sheet_id
        → _struct_sheet.id
  _struct_sheet_range.beg_label_asym_id
        → _struct_asym.id
  _struct_sheet_range.beg_label_comp_id
        → _chem_comp.id
  _struct_sheet_range.beg_label_seq_id
        → _atom_site.label_seq_id
  _struct_sheet_range.end_label_asym_id
        → _struct_asym.id
  _struct_sheet_range.end_label_comp_id
        → _chem_comp.id
  _struct_sheet_range.end_label_seq_id
        → _atom_site.label_seq_id
  _struct_sheet_range.beg_auth_asym_id
        → _atom_site.auth_atom_id
  _struct_sheet_range.beg_auth_comp_id
        → _atom_site.auth_comp_id
  _struct_sheet_range.beg_auth_seq_id
        → _atom_site.auth_seq_id
  _struct_sheet_range.end_auth_asym_id
        → _atom_site.auth_atom_id
  _struct_sheet_range.end_auth_comp_id
        → _atom_site.auth_comp_id
  _struct_sheet_range.end_auth_seq_id
        → _atom_site.auth_seq_id
  _struct_sheet_range.symmetry
```

(*d*) STRUCT_SHEET_ORDER

```
• _struct_sheet_order.range_id_1
        → _struct_sheet_range.id
• _struct_sheet_order.range_id_2
        → _struct_sheet_range.id
• _struct_sheet_order.sheet_id
        → _struct_sheet.id
  _struct_sheet_order.offset
  _struct_sheet_order.sense
```

(*e*) STRUCT_SHEET_HBOND

```
• _struct_sheet_hbond.range_id_1
        → _struct_sheet_range.id
• _struct_sheet_hbond.range_id_2
        → _struct_sheet_range.id
• _struct_sheet_hbond.sheet_id
        → _struct_sheet.id
  _struct_sheet_hbond.range_1_beg_label_atom_id
        → _atom_site.label_atom_id
  _struct_sheet_hbond.range_1_beg_label_seq_id
        → _atom_site.label_seq_id
  _struct_sheet_hbond.range_1_end_label_atom_id
        → _atom_site.label_atom_id
  _struct_sheet_hbond.range_1_end_label_seq_id
        → _atom_site.label_seq_id
  _struct_sheet_hbond.range_2_beg_label_atom_id
        → _atom_site.label_atom_id
  _struct_sheet_hbond.range_2_beg_label_seq_id
        → _atom_site.label_seq_id
  _struct_sheet_hbond.range_2_end_label_atom_id
        → _atom_site.label_atom_id
  _struct_sheet_hbond.range_2_end_label_seq_id
        → _atom_site.label_seq_id
  _struct_sheet_hbond.range_1_beg_auth_atom_id
        → _atom_site.auth_atom_id
  _struct_sheet_hbond.range_1_beg_auth_seq_id
        → _atom_site.auth_seq_id
  _struct_sheet_hbond.range_1_end_auth_atom_id
        → _atom_site.auth_atom_id
  _struct_sheet_hbond.range_1_end_auth_seq_id
        → _atom_site.auth_seq_id
  _struct_sheet_hbond.range_2_beg_auth_atom_id
        → _atom_site.auth_atom_id
  _struct_sheet_hbond.range_2_beg_auth_seq_id
        → _atom_site.auth_seq_id
  _struct_sheet_hbond.range_2_end_auth_atom_id
        → _atom_site.auth_atom_id
  _struct_sheet_hbond.range_2_end_auth_seq_id
        → _atom_site.auth_seq_id
```

The bullet (•) *indicates a category key. Where multiple items within a category are marked with a bullet, they must be taken together to form a compound key. The arrow* (→) *is a reference to a parent data item.*

Different methods of describing β-sheets are in widespread use. The mmCIF dictionary provides data items for two methods and it is anticipated that future versions of the dictionary could cover others. The model used in the STRUCT_SHEET_TOPOLOGY category is the simpler of the two. It is a convenient shorthand for describing the topology, but it does not provide details about strand registration and it is not suitable for describing sheets that contain strands from more than one polypeptide. A more general model is provided by the linked data items in the STRUCT_SHEET_RANGE, STRUCT_SHEET_ORDER and STRUCT_SHEET_HBOND categories. For both methods of representing β-sheets, data items in the parent category STRUCT_SHEET can be used to provide an identifier for each sheet, a free-text description of its type, the number of participating strands and a free-text description of any peculiar aspects of the sheet. The relationships between categories used to describe β-sheets are shown in Fig. 3.6.7.14.

In the description of β-sheet topology based on the STRUCT_SHEET_TOPOLOGY category, the strand that occurs first in the polypeptide chain is numbered 1. Subsequent strands are described by their position in the sheet relative to the previous strand (+1, −3 *etc.*) and by their orientation relative to the previous strand (parallel or antiparallel).

While writing this chapter, a few errors in the mmCIF dictionary were discovered. The use of `_struct_sheet_topology.range_id_1` and `*_2` as pointers to the residues participating in β-sheets is one; the correct data items should be `_struct_sheet_topology.comp_id_1` and `*_2`, and these data items should be pointers to `_atom_site.label_comp_id`. This error will be corrected in future versions of the dictionary. As the data model encoded in the current version of the dictionary is incorrect, no example of its use is given.

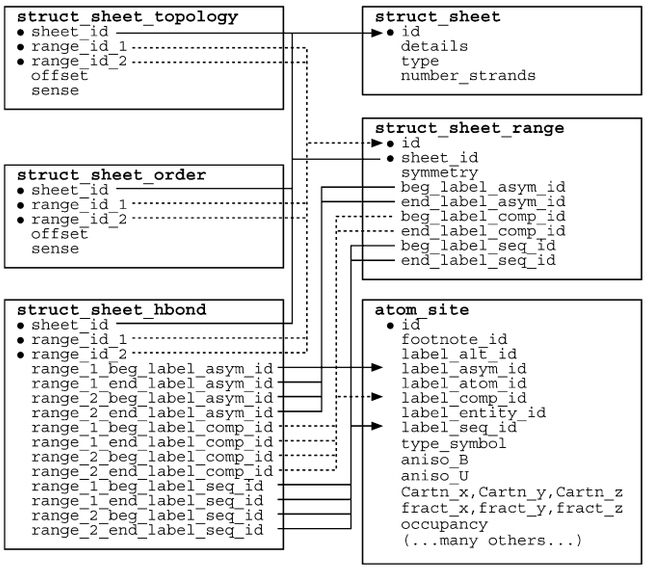

Fig. 3.6.7.14. The family of categories used to describe β-sheets. Boxes surround categories of related data items. Data items that serve as category keys are preceded by a bullet (•). Lines show relationships between linked data items in different categories with arrows pointing at the parent data items.

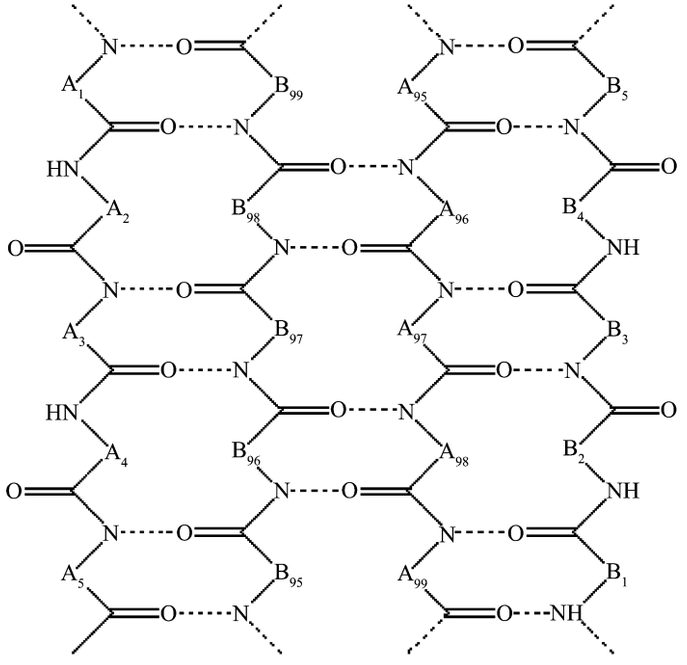

Fig. 3.6.7.15. A hypothetical β-sheet to be described with data items in the STRUCT_SHEET, STRUCT_SHEET_ORDER, STRUCT_SHEET_RANGE and STRUCT_SHEET_HBOND categories. Note that the strands come from two different polypeptides, labelled A and B.

In the more detailed and more general method for describing β-sheets, data items in the STRUCT_SHEET_RANGE category specify the range of residues that form strands in the sheet, data items in the STRUCT_SHEET_ORDER category specify the relative pairwise orientation of strands and data items in the STRUCT_SHEET_HBOND category provide details of specific hydrogen-bonding interactions between strands (see Fig. 3.6.7.15 and Example 3.6.7.14). Note that the specifiers for the strand ranges include the amino acid (**`*_comp_id`** and **`*_seq_id`**), the chain (**`*_asym_id`**) and a symmetry code (**`_struct_sheet_range.symmetry`**). Thus sheets that are composed of strands from more than one polypeptide chain or from polypeptides in more than one asymmetric unit can be described.

Example 3.6.7.14. *A hypothetical β-sheet described with data items in the STRUCT_SHEET, STRUCT_SHEET_ORDER, STRUCT_SHEET_RANGE and STRUCT_SHEET_HBOND categories.*

```
loop_
  _struct_sheet.id
  _struct_sheet.number_strands
    S1   4

loop_
  _struct_sheet_order.sheet_id
  _struct_sheet_order.range_id_1
  _struct_sheet_order.range_id_2
  _struct_sheet_order.sense
    S1 1 2 anti-parallel
    S1 2 3 anti-parallel
    S1 3 4 anti-parallel
    S2 1 2 anti-parallel

loop_
  _struct_sheet_range.sheet_id
  _struct_sheet_range.id
  _struct_sheet_range.beg_label_comp_id
  _struct_sheet_range.beg_label_asym_id
  _struct_sheet_range.beg_label_seq_id
  _struct_sheet_range.end_label_comp_id
  _struct_sheet_range.end_label_asym_id
  _struct_sheet_range.end_label_seq_id
    S1 1 PRO A 1  LEU A 5
    S1 2 CYS B 95 PHE B 99
    S1 3 CYS A 95 PHE A 99
    S1 4 PRO B 1  LEU B 5

loop_
  _struct_sheet_hbond.sheet_id
  _struct_sheet_hbond.range_id_1
  _struct_sheet_hbond.range_id_2
  _struct_sheet_hbond.range_1_beg_label_atom_id
  _struct_sheet_hbond.range_1_beg_label_seq_id
  _struct_sheet_hbond.range_2_beg_label_atom_id
  _struct_sheet_hbond.range_2_beg_label_seq_id
    S1 1 2 A 3  O 97
    S1 2 3 B 98 O 96
    S1 3 4 A 97 O 3
```

It is conventional to assign the number 1 to an outermost strand. The choice of which outermost strand to number as 1 is arbitrary, but would usually be the strand encountered first in the amino-acid sequence. The remaining strands are then numbered sequentially across the sheet.

In some simple cases, the complete hydrogen bonding of the sheet could be inferred from the strand-range pairings and the relationship between the strands (parallel or antiparallel). However, in most cases it is necessary to specify at least one hydrogen bond between adjacent strands in order to establish the registration. The data items in the STRUCT_SHEET_HBOND category can be used to do this. Hydrogen bonds also need to be specified precisely when a sheet contains a nonstandard feature such as a β-bulge. This is a case where it is sufficient to specify a single hydrogen-bonding interaction to establish the registration; here only the **`*_beg_*`** or **`*_end_*`** data items need to be used to reference the atom-label components. However, it is preferable, wherever possible, to specify the initial and final atoms of the two ranges participating in the hydrogen bonding.

3.6.7.5.8. *Molecular sites*

The data items in these categories are as follows:

(*a*) STRUCT_SITE

- **`_struct_site.id`**
  `_struct_site.details`