**Classification and use of topological data**

By V. A. Blatov, D. M. Proserpio and R. M. Hanson

**1. Introduction**

The crystallographic data obtained from experiment or theoretical modeling contain no explicit information on the crystal structure connectivity. However, this information, which might describe all sorts of interatomic bonding as well as the methods of connection of complex structural units, is crucial for many tasks of crystal chemistry and materials science, such as classification, comparative analysis, modeling of crystal growth and assembling, as well as search for correlations among crystal composition, structure and properties. Similar to molecular graphs, which are widespread in organic chemistry, crystal structure can be represented as a three-periodic graph, which is usually called a *net* (Delgado-Friedrichs & O’Keeffe, 2005; Delgado-Friedrichs *et al*., 2005). Unlike molecular graphs, nets are infinite and require special methods of description. Such descriptions have become increasingly common in routine crystal structure determinations and usually follow the conventional crystallographic description, which includes atomic coordinates, space group, unit cell dimensions, and other geometrical information, but does not concern the structure connectivity. Many topological descriptors have been proposed for nets; all of these descriptors are determined from the initial crystallographic data using algorithms. Software implementing these algorithms, such as *Gavrog Systre* (Delgado-Friedrichs & O’Keeffe, 2003) or *ToposPro* (Blatov *et al*., 2014) has made topological analysis available for the general crystallographic community and has already led to the development of special databases (O'Keeffe *et al*., 2008; Blatov *et al*., 2014) that collect the topological parameters of crystal structures. This work will be facilitated by an extension of the CIF format that makes it possible to store topological information along with a crystal structure and to describe the crystal architecture at different levels of organization.

Although the core CIF dictionary contains some topological information in the CHEMICAL\_CONN, GEOM\_BOND, and GEOM\_CONTACT categories, this information is limited and does not cover all topological aspects of a crystal. The core CIF dictionary does not have the ability to describe a plethora of non-valence intra- or intermolecular interactions (hydrogen bonds, halogen and other specific bonds, van der Waals interactions, *etc*.). Moreover, structures that contain polyatomic *building units*, such as molecules, metal complex groups, clusters, *etc*., can be simplified to underlying nets, which bear the general topological motif of these units, and these nets should also be formalized in an appropriate format (Alexandrov *et al*. 2011; Bonneau *et al.* 2018; Barthel *et al.* 2018; Shevchenko & Blatov, 2021). The topoCIF dictionary solves all of these issues, making the exchange of the topological information easier.

**2. Dictionary design considerations**

**2.1 Overall format** The topoCIF dictionary can be expressed in either CIF 1 or CIF 2 format. Where there is a difference, it is in the use of CIF 2 lists for just a few data items. The discussion below lists the CIF 2 data keys; ⱡ indicates the availability of CIF 1 equivalents, which are discussed where appropriate.

**2.2 Categories and scope**

The topoCIF dictionary contains categories for the description of connectivity and topological properties of periodic nets. Since the corresponding terminology is still under development, only conventional terms and descriptors defined by Delgado-Friedrichs & O’Keeffe (2005) and Blatov *et al*. (2010) are included. The TOPOL\_ENTANGL category has been left empty until the terminology on entanglements gains consistency. The categories can be divided into two groups: (i) descriptors of crystal structure connectivity, and (ii) descriptors of topological properties of a periodic net. The descriptors of the first group can be computed directly from the initial crystallographic data; the descriptors of the second group are determined based on that connectivity information. Thus, the topoCIF dictionary enables one to generate two new levels of the crystal structure description: a connected net without any further detailing and a topological description of that connected net.

**3. Classification of data definitions**

**3.1. Underlying net description**

The *underlying net* is the net of centroids of structural units. The edges of the net represent the links between these units. The underlying net describes the topology (Francl, 2008) of a particular representation of a crystal structure (Eon, 2016). This representation is said to be *complete* if the underlying net coincides with the initial structure (that is, maps all atoms in the ATOM\_SITE block), and partial if not. For example, diamond and SiO2 (cristobalite) can be represented by the same underlying net **dia**. For diamond, this net is complete, because all the bonding we would normally associate with the diamond structure are represented by the net. But for cristobalite, **dia** is only a partial net, since it only represents the silicon atoms. The oxygen atoms can be considered to be bridges (links) between silicon atoms, but those links are represented just by edges of the **dia** net. In the discussion below, bullet (∙) indicates the category key, typically an integer, and arrow (→) indicates a reference to another category’s key having the same value.

All topological data in the TOPOL category describe one or several possible (complete or partial) topological representations of the crystal structure given in the form of a periodic net. The nodes of this net either coincide with the atoms or correspond to the centers of building units. The main two categories, TOPOL\_LINK and TOPOL\_NODE, describe the underlying net and its connectivity. The structure can often consist of several interpenetrating nets, which are not interconnected. Each symmetry independent isolated net must be described separately in the TOPOL\_NET category (see part 3.2), and labeled by a \_topol\_net.id identifier. The relations between the atoms, which are listed in the ATOM\_SITE block, and nodes and links of the underlying net are established in the TOPOL\_ATOM category.

Scheme I shows the overall relationship among TOPOL\_NET, TOPOL\_LINK, TOPOL\_NODE, and TOPOL\_ATOM. The key category is TOPOL\_LINK, with references to TOPOL\_NET and TOPOL\_NODE. TOPOL\_ATOM provides coordinates for atoms associated with nodes and links. All three categories can reference ATOM\_SITE directly, depending upon the complexity of the net.



TOPOL\_LINK

A basic net that only references nodes that are single crystallographic atoms requires only the TOPOL\_LINK category.

* \_topol\_link.id

\_topol\_link.atom\_label\_1 → \_atom\_site.label

\_topol\_link.atom\_label\_2 → \_atom\_site.label

\_topol\_link.symop\_1 → \_space\_group\_symop.id

\_topol\_link.translation\_1ⱡ

\_topol\_link.symop\_2 → \_space\_group\_symop.id

\_topol\_link. translation\_2ⱡ

\_topol\_link .id is the unique category key, an integer, typically starting with 1. It is only required if TOPOL\_ATOM is present, as that is where it is referenced. Similar to the GEOM\_BOND category, TOPOL\_LINK allows for a general reference to an ATOM\_SITE atom along with a symmetry operation and a translation to be applied to that atom’s reference position. The syntax, however, is slightly different. Like \_geom\_bond.symmetry\_\*, \_topol\_link.symop\_\* references \_space\_group\_symop.id, but the added translational component is indicated as a list [ i j k ] with [1 0 0] representing a unit translation along the *a* axis. (This allows for a more general representation than the “555” notation used in \_space\_gorup\_symop.id.) Both the identity operation (*x,y,z*, always 1), and no translation, [0 0 0], can be represent using the CIF “default” indicator, an unquoted full stop (period) character. The CIF 1 equivalents for \_topol\_link.translation\_\* are \_topol\_link.translation\_\*\_x, \_topol\_link.translation\_\*\_y, and \_topol\_link.translation\_\*\_z, where \* is 1 or 2. Thus, the diamond net can be described simply with the following set of tags (Example 1):

loop\_

\_topol\_link.atom\_label\_1

\_topol\_link.atom\_label\_2

\_topol\_link.symop\_1

\_topol\_link.translation\_1

\_topol\_link.symop\_2

\_topol\_link.translation\_2

C1 C1 1 . 13 .

where the only data record in atom\_site is for atom\_site.label = C1.

Additional items describing links include:

\_topol\_link.distance

\_topol\_link.label

\_topol\_ link.net\_id → \_topol\_net.id

\_topol\_link.multiplicity

\_topol\_link.node\_id\_1 → \_topol\_node.id

\_topol\_link.node\_id\_2 → \_topol\_node.id

\_topol\_link.order

\_topol\_link.special\_details

\_topol\_link.type

\_topol\_link.Voronoi\_solidangle

Of these data items, only \_topol\_link.net\_id is ever required, and only then if the TOPOL\_NET category is present and there is more than one net. Note that any number of links may correspond with a given net, but each link must correlate with only one net. When the TOPOL\_NODE category is present, \_topol\_link.node\_id\_1 and/or \_topol\_link.node\_id\_2 point to the data item in that category associated with the two ends of the link. These data items are described in the next section.

The data items \_topol\_link.distance, \_topol\_link.multiplicity, \_topol\_link.order, \_topol\_link.special\_details, \_topol\_link.type, and \_topol\_link.Voronoi\_solidangle are additional optional data items that relate characteristics of the link. For example, \_topol\_link.Voronoi\_solidangle , designates the solid angle of an interatomic contact A-X, which corresponds to a common face of the Voronoi polyhedra of A and X atoms (Blatov, 2004; Fig. 1). These items are described more fully in the dictionary descriptions.



**Figure 1:** The solid angle of the shaded Voronoi polyhedron face in the body-centered cubic lattice is equal to the shaded segment of the unit sphere being cut off by the pyramid with the Voronoi polyhedron atom A at the vertex and the face in the base. The shaded face is shared between Voronoi polyhedra of A and X atoms and corresponds to the bond A-X.

|  |
| --- |
| *Example 1. Connectivity of the diamond crystal structure. All atoms coincide with the nodes and all bonds coincide with the edges, so the atomic network coincides with the underlying net. Thus, no special TOPOL\_NET and TOPOL\_NODE sections are needed.*  loop\_  \_space\_group\_symop.id  \_space\_group\_symop.operation\_xyz  1 x,y,z  2 1/4-x,1/4-y,z  # Symmetry operations skipped  13 -y,-x,-z  # Symmetry operations skipped  192 3/4-z,1/2+y,1/4-x  loop\_  \_atom\_site.label  \_atom\_site.type\_symbol  \_atom\_site.symmetry\_multiplicity  \_atom\_site.fract\_x  \_atom\_site.fract\_y  \_atom\_site.fract\_z  \_atom\_site.occupancy  C1 C 8 0.12500 0.12500 0.12500 1.0000  \_topol\_net.overall\_topology\_RCSR dia  loop\_  \_topol\_link.id  \_topol\_link.atom\_label\_1  \_topol\_link.atom\_label\_2  \_topol\_link.symop\_1  \_topol\_link.translation\_1  \_topol\_link.symop\_2  \_topol\_link.translation\_2  \_topol\_link.distance  \_topol\_link.Voronoi\_solidangle  \_topol\_link.type  \_topol\_link.multiplicity  1 C1 C1 1 [0 0 0] 13 [0 0 0] 1.5446 22.04 v 16 |

TOPOL\_NODE

A succinct alternative to directly describing the atoms and symmetry in the TOPOL\_LINK category is to move that information to a TOPOL\_NODE loop and simply point to nodes involved in the link. This provides a better separation between the mathematical abstraction of a net of links and nodes and the geometry associated with that net and is preferred when the two ends of a link refer to symmetry-equivalent nodes. Thus, the above diamond example could be described as follows, where the TOPOL\_LINK loop reduces to a simple n x 2 matrix:

loop\_

\_topol\_node.id

\_topol\_node.atom\_label

\_topol\_node.symop

\_topol\_node.translation

1 C1 1 .

2 C1 13 .

loop\_

\_topol\_link.node\_id\_1

\_topol\_link.node\_id\_2

1 2

Depending upon the circumstances, symmetry-related items may be left in TOPOL\_LINK:

loop\_

\_topol\_node.id

\_topol\_node.atom\_label

1 C1

loop\_

\_topol\_link.node\_id\_1

\_topol\_link.node\_id\_2

\_topol\_link.symop\_1

\_topol\_link.translation\_1

\_topol\_link.symop\_2

\_topol\_link.translation\_2

1 1 . 1 13 .

Thus, the TOPOL\_LINK and TOPOL\_NODE categories provide a very flexible format for the representation of periodic nets.

More generally, the TOPOL\_NODE category involves the following essential keys:

\_topol\_node.atom\_label → \_atom\_site\_label

\_topol\_node.fract\_x

\_topol\_node.fract\_y

\_topol\_node.fract\_z

* \_topol\_node.id

\_topol\_node.symop

\_topol\_node.translationⱡ

\_topol\_node.id is a required tag, as it is the category key, referenced by TOPOL\_LINK and TOPOL\_ATOM. Its integer value must be unique and typically the series starts with the value 1. \_topol\_node.atom\_label is required only if one or more of the nodes corresponds to single crystallographic atoms and the TOPOL\_ATOM category is not present. Otherwise, that reference can be moved to TOPOL\_ATOM (see below). The data items \_topol\_node.fract\_x, \_topol\_node.fract\_y, and \_topol\_node.fract\_z are symmetry-transformed fractional coordinates. If present for the loop, they should be listed as the default ‘.’ whenever the symmetry transformation is the default transformation (x,y,z [0 0 0], explicitly or implied) and the node is not referenced from TOPOL\_ATOM. When fractional coordinates are present for a packet in \_topol\_node, \_topol\_node.symop and \_topol\_node.translation should not be present (or, if present, should match the given transformed fractional coordinates).

Additional TOPOL\_NODE descriptors include:

\_topol\_node.chemical\_formula\_iupac

\_topol\_node.chemical\_formula\_moiety

\_topol\_node.chemical\_formula\_sum

\_topol\_node.coordination\_sequence

\_topol\_node.coordination\_sequence\_plain

\_topol\_node.extended\_point\_symbol

\_topol\_node.label

\_topol\_node.net\_id → \_topol\_net.id

\_topol\_node.point\_symbol

\_topol\_node.structural\_formula\_InChI

\_topol\_node.structural\_formula\_SMILES

\_topol\_node.symmetry\_multiplicity

\_topol\_node.vertex\_symbol

\_topol\_node.Wyckoff\_symbol

Of these data items, only \_topol\_node.net\_id is ever required, and only then if the TOPOL\_NET category is present and there is more than one net. Note that, as for links, any number of nodes may correspond with a given net, but each node must correlate with only one net. As for fractional coordinates, for nodes with single associated atoms from ATOM\_SITE, \_topol\_node.symmetry\_multiplicity and \_topol\_node.Wyckoff\_symbol should not duplicate information in the ATOM\_SITE category.

TOPOL\_ATOM

When some of the nodes of a net do not correspond with an atom of the crystal structure, or links themselves are abstractions of atoms or groups of atoms (*building units*, Alexandrov *et al*. 2011; Bonneau *et al.* 2018; Shevchenko & Blatov, 2021) the TOPOL\_ATOM category is required along with TOPOL\_NODE and TOPOL\_LINK. These three categories, then, describe the full chemical constitution and connectivity of the underlying net. The TOPOL\_ATOM category is quite simple:

\_topol\_atom.atom\_label

\_topol\_atom.element\_symbol

* \_topol\_atom.id

\_topol\_atom.link\_id → \_topol\_link.id

\_topol\_atom.node\_id → \_topol\_node.id

\_topol\_atom.symop

\_topol\_atom.translationⱡ

\_topol\_atom.fract\_x

\_topol\_atom.fract\_y

\_topol\_atom.fract\_z

Of these items, only \_topol\_atom.atom\_label is absolutely required. Though \_topol\_atom.id is a key, within the current specification it is not referenced. Exactly one of \_topol\_atom.node\_id or \_topol\_atom.link\_id is required, pointing to the node or link that this atom is associated with. Note that this allows any number of atoms to be associated with a given node or link, but not to more than one node or link. \_topol\_atom.element\_symbol is strongly recommended, as there is no other location in a CIF file where a specific element symbol is required. In the case that for some reason the atom does not relate to a specific element (due to occupational disorder or some other reason), this field should be left as unquoted ‘.’. The relationship between symmetry transformation and fractional coordinates is the same as for TOPOL\_NODE. Fractional coordinates are symmetry-transformed coordinates, and, if both the symmetry transformation and factional coordinates are present and non-default values, they must match.

Using TOPOL\_ATOM, the diamond example could be described as follows:

loop\_

\_topol\_atom.atom\_label

\_topol\_atom.symop

\_topol\_atom.translation

C1 1 .

C1 13 .

loop\_

\_topol\_node.id

1

2

loop\_

\_topol\_link.node\_id\_1

\_topol\_link.node\_id\_2

1 2

For such a simple case, the use of TOPOL\_ATOM and TOPOL\_NODE are clearly unnecessary. But consider the following pair of nets (Example 2):

         loop\_

         \_topol\_net.id

         \_topol\_net.label

         \_topol\_net.special\_details

         \_topol\_net.overall\_topology\_TOPOS

         1 Net\_1 'Atomic net' 'Unknown'

         2 Net\_2 'Underlying net with carbonyl ligands as nodes' '2,4T3'

We describe the links as a simple topological matrix, with additional descriptive information:

         loop\_

         \_topol\_link.id

         \_topol\_link.net\_id

         \_topol\_link.node\_id\_1

         \_topol\_link.node\_id\_2

         \_topol\_link.distance

         \_topol\_link.type

         \_topol\_link.multiplicity

         1 1 1 2   1.9121 v 4  # Li1 - O1

         2 1 2 3   1.1452 v 4  # O1 - C1

         3 1 2 4   1.7422 v 4  # C1 - Co1

         4 2 5 6   2.4032 v 4  # Li1 - CO

         5 2 6 7   2.3963 v 4  # CO - Co

Adding the nodes, with labels and chemical formulas:

         loop\_

         \_topol\_node.id

         \_topol\_node.net\_id

         \_topol\_node.label

         \_topol\_node.chemical\_formula\_sum

         1 1 Li1 Li

2 1 O1  O

         3 1 C1  C

         4 1 Co1 Co

         5 2 ZA1 Li

         6 2 ZB1 CO

         7 2 ZC1 Co

Note that Node 6 has two atoms associated with, the carbon and oxygen atoms of a carbonyl group. Finally, specifying the positions of those nodes that have special labels:

loop\_

\_topol\_atom.id

\_topol\_atom.node\_id

\_topol\_atom.label

\_topol\_atom.chemical\_formula\_sum

1 5 Li1  Li

2 6 C1   C

3 6 O1   O

4 7 Co1  Co

For a much more involved example, see the metal-organic framework MOF5 example at GitHub (<https://github.com/COMCIFS/TopoCif/blob/master/more_examples/mof5-v2d-three_nets.cif>), involving three distinct nets, ten links, eleven nodes, and sixty atoms.

|  |
| --- |
| *Example 2. Connectivity of atomic and underlying nets for an interpenetrating array of two LiCo(CO)4 networks. The atomic net consists of Li, C, O, and Co atoms, while the underlying net is built from three kinds of nodes: Li and Co atoms and carbonyl (CO) ligand; the nodes are labeled as ZA1, ZC1, and ZB1, respectively. The \_topol\_node\_\* items include references to atom labels for the atoms and coordinates for the nodes. Some fields, which values are not required or should be taken from the ATOM\_SITE block, are specified with the '.' symbol. Two possible variants are shown: the coordinates of ZA1 are specified by a reference to Li1 atom, while the coordinates of ZC1 are specified explicitly. Both atomic and underlying nets are described in the TOPOL\_NET section.*  loop\_  \_space\_group\_symop.id  \_space\_group\_symop.operation\_xyz  1 x,y,z  2 -x,-y,z  3 x,-y,-z  # Symmetry operations skipped  24 -z,y,-x  loop\_  \_atom\_site.label  \_atom\_site.type\_symbol  \_atom\_site.site\_symmetry\_multiplicity  \_atom\_site.fract\_x  \_atom\_site.fract\_y  \_atom\_site.fract\_z  \_atom\_site.occupancy  Li1 Li 1 0.00000 0.00000 0.00000 1.0000  C1 C 4 0.31850 0.31850 0.31850 1.0000  O1 O 4 0.19920 0.19920 0.19920 1.0000  Co1 Co 1 0.50000 0.50000 0.50000 1.0000  loop\_  \_topol\_net\_id  \_topol\_net.special\_details  \_topol\_net.overall\_topology\_TOPOS  \_topol\_net.z\_number  Net\_1 'Atomic net' 'Unknown' 2  Net\_2 'Underlying net with carbonyl ligands as nodes' '2,4T3' 2  loop\_  \_topol\_node.id  \_topol\_node.label  \_topol\_node.net\_id  \_topol\_node.chemical\_formula\_sum  \_topol\_node.atom\_label  \_topol\_node.fract\_x  \_topol\_node.fract\_y  \_topol\_node.fract\_z  1 Li1 Net\_1 . Li1 . . .  2 C1 Net\_1 . C1 . . .  3 O1 Net\_1 . O1 . . .  4 Co1 Net\_1 . Co1 . . .  5 ZA1 Net\_2 Li . 0.00000 0.00000 0.00000  6 ZB1 Net\_2 CO . 0.25036 0.25036 0.25036  7 ZC1 Net\_2 Co . 0.50000 0.50000 0.50000  loop\_  \_topol\_atom.node\_id  \_topol\_atom.atom\_label  6 C1  6 O1  loop\_  \_topol\_link.id  \_topol\_link.node\_id\_1  \_topol\_link.node\_id\_2  \_topol\_link.distance  \_topol\_link.symop\_1  \_topol\_link.translation\_1  \_topol\_link.symop\_2  \_topol\_link.translation\_2  \_topol\_link.type  \_topol\_link.multiplicity  1 1 3 1.9121 1 [0 0 0] 3 [0 0 0] v 4  2 2 3 1.1452 1 [0 0 0] 1 [0 0 0] v 4  3 2 4 1.7422 1 [0 0 0] 1 [0 0 0] v 4  4 5 6 2.4032 1 [0 0 0] 3 [0 0 0] v 4  5 6 7 2.3963 1 [0 0 0] 1 [0 0 0] v 4 |

**3.2. Underlying net topological properties**

The topological properties of underlying nets described by the TOPOL\_LINK, TOPOL\_NODE and TOPOL\_ATOM categories are collected in the remaining TOPOL categories, namely TOPOL\_NET, TOPOL\_ENTANGL, TOPOL\_TILING (including TOPOL\_TILING\_FACES and TOPOL\_TILING\_TILE), and TOPOL\_OCCURRENCE categories. These categories are discussed below.

TOPOL\_NET

The TOPOL\_NET category enables one to describe a crystal structure in different representations, which depend on the selected structural units and levels of interatomic interaction (Alexandrov *et al*., 2011). This approach can resolve the problem of ambiguity in the description of the crystal structure connectivity: the crystal structure can be treated in different ways and described by several underlying nets listed in the TOPOL\_NET loop of a data block. If several nets are described, or additional net-level descriptions are needed, the TOPOL\_NET category is necessary. Its items include:

\_topol\_net.genus

* \_topol\_net.id

\_topol\_net.name

\_topol\_net.occurrence\_total

\_topol\_net.overall\_topology\_EPINET

\_topol\_net.overall\_topology\_IZA

\_topol\_net.overall\_topology\_RCSR

\_topol\_net.overall\_topology\_SP

\_topol\_net.overall\_topology\_TOPOS

\_topol\_net.period

\_topol\_net.special\_details

\_topol\_net.td10

\_topol\_net.total\_point\_symbol

\_topol\_net.z\_number

\_topol\_net.id, again a serial integer, is required if there is more than one net, as it is the key referenced from TOPOL\_LINK and TOPOL\_NODE. \_topol\_net.name allows for an optional arbitrary name for the net. The overall topology of the net can be designated in accordance with five mostly recommended nomenclatures (RCSR, EPINET, SP, TOPOS or IZA), which have special tags:

* RCSR lower-case bold three-letter symbols, see O'Keeffe *et al*. (2008) and http://rcsr.net/ for details. Example: **dia** (see http://rcsr.net/nets/dia)
* EPINET *sqc*XXXXX symbols, see Ramsden *et al*. (2009) and http://epinet.anu.edu.au/ for details.
* SP (sphere packing) symbols *k*/*m*/*fn* as proposed by Koch *et al*. (2006).
* TOPOS symbols N*D*n (Aman *et al*., 2014), where N is a sequence of degrees (coordination numbers) of all independent nodes; *D* is one of the letters *C*, *L*, or *T* designating the dimensionality of the net (*C* – chain, *L* – layer, *T* – three-periodic); n enumerates non-isomorphic nets with a given N*D* sequence. For instance, the symbol 3,3,4*T*3 denotes the 3rd (by the order) three-periodic trinodal net with two 3-coordinated and one 4-coordinated independent nodes. For finite (molecular) graphs the symbols N*M*k-n are used, where k is the number of vertices (atoms) in the graph.
* IZA-SC capital three-letter symbols for zeolites, see http://www.iza-structure.org/databases/ for details.

If the topology is designated in another nomenclature it should be described in \_topol\_net.special\_details, where an arbitrarily detailed description of the net can be provided.

Additionally, this category contains conventional topological descriptors of a net, such as net periodicity (\_topol\_net.period), topological density TD10 (\_topol\_net.td10), point symbol (\_topol\_net.total\_point\_symbol) and number of nets (*Z*) in the structure (\_topol\_net.z\_number). For more details and definitions see Delgado-Friedrichs & O’Keeffe (2005), Delgado-Friedrichs *et al*. (2005) and Blatov *et al*. (2010). \_topol\_net.occurrence\_total specifies the total occurrence of the net topology in crystal structures as reported in the literature or databases (e.g. TopCryst; https://topcryst.com).

TOPOL\_ENTANGL

This category contains no items, because the descriptors of entanglements in crystal structures have not been standardized yet. However, the importance of entanglements in the topological description is acknowledged , and this category is expected to be developed in future versions of the dictionary.

TOPOL\_TILING, TOPOL\_TILING\_FACES, and TOPOL\_TILING\_TILE

These three categories describe a tiling, which is carried by the underlying net. Since only a single unentangled net can carry tilings, these data can be included only if the structure representation is described by a single net. However, if several representations are described and hence there are several lines in the TOPOL\_NET block, each underlying net can have its own tiling described in the TOPOL\_TILING block. As a rule, natural tiling (Blatov *et al*., 2007) is described, since it contains minimal cages, which can be combined to provide all other larger cages, including infinite channels.

Items in these categories include:

\_topol\_tiling.Dsize

\_topol\_tiling.dual

\_topol\_tiling.edges

\_topol\_tiling.faces

\_topol\_tiling.signature

\_topol\_tiling.tiles

\_topol\_tiling.vertices

\_topol\_tiling\_faces.tile\_id → \_topol\_tiling\_tile.id

\_topol\_tiling\_faces.size

\_topol\_tiling\_tile.count

* \_topol\_tiling\_tile.id

TOPOL\_OCCURRENCE

The TOPOL\_OCCURRENCE category stores information on the occurrences of a particular net topology in other crystal structures and can be used together with the \_topol\_net.overall\_topology\_\* items for developing topological databases. Items include:

\_topol\_occurrence.citation\_id → \_citation.id

* \_topol\_occurrence.id

\_topol\_occurrence.net\_id → \_topol\_net.id

Reference or collection codes of crystal structure determinations can be used as \_topol\_occurrence.id values. For example, reference codes may return to items in the Cambridge Structural Database or the Inorganic Crystal Structure Database (Example 3).

|  |
| --- |
| *Example 3. Description of the occurrence of a particular topology specified in the TOPOL\_NET category. Here all occurrences refer to structural determinations from the CSD. The total occurrence is specified in \_topol\_net.occurrence\_total; actually, it includes all crystal structures with this topology that are currently known.*  loop\_  \_topol\_net\_id  \_topol\_net.special\_details  \_topol\_net.overall\_topology\_RCSR  \_topol\_net.occurrence\_total  1 'Occurrence data are taken from topcryst.com' dia-a-a 5  loop\_  \_citation\_id  \_citation\_database\_id\_CSD  1 MUMYIC  2 MUNDAA  3 MUMYUO  4 MUMYEY  5 MUMYAU  loop\_  \_topol\_occurrence.id  \_topol\_occurrence.net\_id  \_topol\_occurrence.citation\_id  1 1 1  2 1 2  3 1 3  4 1 4  5 1 5 |

**3.3. Development of the supporting software and databases**

The CIF topological data can be read and generated by the *ToposPro* program package (Blatov *et al*., 2014) and online service *TopCryst* (https://topcryst.com). *Jmol*/*JSmol* software (<https://jmol.sourceforge.net>, <https://github.com/BobHanson/Jmol-SwingJS>) can also read and visualize the nets specified in the topoCIF format. The topological databases Reticular Chemistry Structure Resource (RCSR, http://rcsr.net/) and ToposPro Collections (TTD and TTO, https://topospro.com) use the topoCIF topological descriptors. Thus, one can use the topoCIF format for exporting or importing information to or from these databases.

**4. Net Reconstruction from CIF Data**

The periodic net described by a topoCIF file is fully defined by its set of TOPOL\_ATOM, TOPOL\_NODE, and TOPOL\_LINK data. The method of constructing a net for visualization or analysis from this data is not trivial, since generally a CIF file only provides locations for the asymmetric unit. Below we describe a sketch of the sequence of steps used by Jmol. The presumption is that we start with a request for the visualization of a net based on a given set of unit cells or a specified volume of space to be filled. Net construction proceeds in the following order:

* Reading relevant \_cell, \_symmetry, \_symmetry\_equiv\_pos, \_atom\_site, \_topol\_net, \_topol\_link, \_topol\_node, and \_topol\_atom data (which could come in any order within a CIF file) from a data block.
* Processing all the TOPOL\_ATOM items by
  + applying indicated symmetry operations and translations, and
  + associating each atom with its designated link or node.
* Creating any additional nodes needed for links that have not been already defined from TOPOL\_ATOM, including node-specific symmetry operations and translations.
* Processing all additional link-specific symmetry operations and translations.
* Applying all space group symmetry operations for all nodes and atoms associated with links or polyatomic nodes.
* Removing any unconnected or unassociated nodes that are present within the specified space.

A reference dual Java/JavaScript implementation can be found at SourceForge (<https://sourceforge.net/p/jmol/code/HEAD/tree/trunk/Jmol/src/org/jmol/adapter/readers/cif/TopoCifParser.java>) or GitHub (<https://github.com/BobHanson/Jmol-SwingJS/blob/master/src/org/jmol/adapter/readers/cif/TopoCifParser.java>).

**References**

Alexandrov, E. V., Blatov, V. A., Kochetkov, A. V. & Proserpio, D. M. (2011). *Underlying nets in three-periodic coordination polymers: topology, taxonomy and prediction from a computer-aided analysis of the Cambridge Structural Database. CrystEngComm*, **13**, 3947-3958.

Aman, F., Asiri, A. M., Siddiqui, W. A., Arshad, M. N., Ashraf, A., Zakharov, N. S. & Blatov, V. A. (2014). *Multilevel topological description of molecular packings in 1,2-benzothiazines. CrystEngComm*, **16**, 1963-1970.

Barthel, S., Alexandrov, E. V., Proserpio, D. M., Smit, B. (2018). *Distinguishing Metal-Organic Frameworks. Cryst. Growth Des.*, **18**, 1738–1747.

Blatov, V.A. (2004). *Voronoi-Dirichlet polyhedra in crystal chemistry: theory and applications. Cryst. Rev.* **10**, 249-318.

Blatov, V. A. (2007). *Topological relations between three-dimensional periodic nets. I. Uninodal nets. Acta Cryst.* **A63**, 329–343.

Blatov, V. A., Delgado-Friedrichs, O., O’Keeffe, M. & Proserpio, D. M. (2007). *Three-periodic nets and tilings: natural tilings for nets. Acta Cryst*. **A63**, 418-425.

Blatov, V.A., O’Keeffe, M. & Proserpio, D. M. (2010). *Vertex-, face-, point-, Schläfli-, and Delaney-symbols in nets, polyhedra and tilings: recommended terminology*. *CrystEngComm*, **12**, 44-48.

Blatov, V. A., & Proserpio, D. M. (2009). *Topological relations between three-dimensional periodic nets. I. Binodal nets. Acta Cryst.* **A65**, 202-212.

Blatov, V.A., Shevchenko, A.P. & Proserpio, D.M. (2014). *Applied topological analysis of crystal structures with the program package ToposPro. Cryst. Growth Des.* **14**, 3576-3586.

Bonneau, C., O’Keeffe, M., Proserpio, D. M., Blatov, V. A., Batten, S. R., Bourne, S. A., Lah, M. S., Eon, J.-G., Hyde, S. T., Wiggin, S. B. & Öhrström, L. (2018) *Deconstruction of Crystalline Networks into Underlying Nets: Relevance for Terminology Guidelines and Crystallographic Databases. Cryst. Growth Des.*, **18**, 3411–3418

Delgado-Friedrichs, O. & O’Keeffe, M. (2003). *Identification of and symmetry computation for crystal nets. Acta Cryst.* **A59**, 351-360.

Delgado-Friedrichs, O. & O’Keeffe, M. (2005). *Crystal nets as graphs: terminology and definitions. J. Solid State Chem*. **178**, 2480-2485.

Delgado-Friedrichs, O., Foster, M. D., O'Keeffe, M., Proserpio, D. M., Treacy, M. M. J. & Yaghi, O. M. (2005). *What do we know about nets? J. Solid State Chem*. **178**, 2533-2554.

Eon, J-.G (2016) *Topological features in crystal structures: a quotient graph assisted analysis of underlying nets and their embeddings.* *Acta Cryst.* **A72**, 268-293.

Francl, M. (2009) *Stretching topology. Nat. Chem.*, **1**, 334-335.

O'Keeffe, M., Peskov, M. A., Ramsden, S. J. & Yaghi, O. M. (2008). *The reticular chemistry structure resource (RCSR) database of, and symbols for, crystal nets. Acc. Chem. Res.* **41**, 1782-1789.

Koch, E., Fischer, W. & Sowa, H. (2006). *Interpenetration of homogeneous sphere packings and of two-periodic layers of spheres. Acta Cryst.* **A62**, 152-167.

Ramsden, S. J., Robins, V., Hyde, S. T. (2009). *Three-dimensional Euclidean nets from two-dimensional hyperbolic tilings: kaleidoscopic examples. Acta Cryst.* **A65**, 81–108.

Shevchenko, A.P. & Blatov, V.A. (2021). *Simplify to understand: how to elucidate crystal structures? Struct. Chem.* **32**, 507-519.