American Mineralogist

CIF Guide

American Mineralogist Technical Editing Team
August 2020

Contents

1. Introduction

The purpose of this document is to guide the author through the process of creation and validation of a CIF (crystallographic information file) for submission to *American Mineralogist.*

The CIF format is the *de facto* standard for communicating all sorts of crystallographic information. In particular it is used to archive crystal structure data, combined with experimental and refinement details.

Manuscripts reporting results of crystal structure refinements or simulations of a complete structure must be accompanied by a CIF providing the structural data. To ensure quality and integrity of the crystallographic data, the submitted CIF will be reviewed by the American Mineralogist technical editing team. After publication of the manuscript, the CIFs will be made available as supplementary material on the American Mineralogist web pages. They may also be deposited into databases, such as the AMCSD (American Mineralogist Crystal Structure Database) and the COD (Crystallography Open Database).

Tables. Since a CIF contains atomic fractional coordinates and displacement parameters as well as sample measurement conditions, separate tables listing these parameters are generally unnecessary in the manuscript and should not be submitted as supplementary material. Exceptions include brief tables of crystallographic parameters (such as selected bond lengths or bond valence sums) that are central to the discussion in the paper, especially if the tables help the reader to understand the comparison of multiple structures.

Only one CIF should be submitted with a manuscript. If more than one structure refinement is reported in a manuscript, each structure refinement must be a separate data block in the CIF. Do not compress or otherwise modify the CIF for submission. The filename extension must be `.cif`.

2. Creating a CIF

Most modern diffractometer control software packages, data reduction software, and structure refinement programs will produce a CIF for you. You are strongly recommended to use these facilities to avoid manual editing of crystallographic data, and thus avoid inevitable 'cut and paste' errors. The recommended sequence of tasks for producing a CIF is:

1. Perform data collection and reduction, and then structure refinement
2. Collate the CIF fragments produced in each step to one cif for one structure, using a CIF-compliant editor.
3. Validate the CIF at the IUCr [3]
4. If more than one structure refinement is being reported, join the individual cifs together into a master CIF (section 6.3), and lock it against any editing
5. Create the data tables and figures for the manuscript from the data in the master CIF.
6. Submit the manuscript and the CIF (not before a final syntax check!).

As the crystal structure data is part of the manuscript, the CIF should be edited and checked with the same care as devoted to the manuscript!

3. CIF Resources

The following are all tools that we find useful in preparing, editing and validating CIFs. The most important is [3], a tool for validating both the syntax and the content of the CIF. Your CIF must pass the basic validation tests at [3] before it can be submitted with a manuscript for publication.

[1] IUCr cif resources: http://www.iucr.org/resources/cif

[2] IUCr Cif Guide: http://journals.iucr.org/b/services/cifguide.html

[3] IUCr online checkCif for validation: http://journals.iucr.org/services/cif/checking/checkbasic.html

[4] enCIFer to create and edit CIFs: http://www.ccdc.cam.ac.uk/Solutions/FreeSoftware/Pages/EnCIFer.aspx

[5] publCIF, to edit CIFs: http://journals.iucr.org/services/cif/publcif

[6] Program to remove illegal characters from a CIF: www.rossangel.net

[7] Program to tabulate data from multiple data blocks in one cif (for example from variable P or T studies) www.rossangel.net

[8] Example CIFs for American Mineralogist: http://www.minsocam.org/MSA/AmMin/CIF_info.html

4. Minimum CIF Requirements

This section defines the requirements for CIFs submitted to American Mineralogist.

A CIF should represent the results of a structural study which is a physically-reasonable crystal structure model, including all relevant information derived from various experimental and/or theoretical sources. In this context 'physically reasonable' normally includes the following:

- crystal-chemically reasonable bond lengths and displacement parameters.
- charge neutral composition in the refined crystal structure model, except when light-atom positions such as H cannot be meaningfully modelled.
- chemical composition: all significant elements should be present in the structure model, with minor elements approximated appropriately. See sections 4.2, 6.4 and 6.5.

All of the items listed in the following sub-sections (4.n) are required items. The values of all experimentally derived parameters must be reported in the CIF, and all must be reported with esd's!

An example file `minimal_cif_example.cif` illustrates these basic requirements. You can find it at http://www.minsocam.org/MSA/AmMin/CIF_info.html.

Hints as to how to achieve these guidelines are given in section 2 above. If you have more complex cases see section 6, below.

4.1 Syntax

The cif file has to meet the cif-syntax as defined by the IUCr [1]. This includes the restriction on line lengths to 80 characters, and the use of only the restricted ASCII character set. See section 6 if you have problems.

Please use available software to perform the editing and cif syntax check [3-5] and to correct errors prior to submission.

For studies reporting multiple structures, please include all structures into a single CIF using a separate data-block for each structure, and appropriate data-block names. See section 6.3 for how to do this.

4.2 Chemical data

following data items are required

```
_chemical_formula_sum              'Si O2'
_chemical_formula_weight           60.08
```

The `_chemical_formula_sum` should reflect the idealised composition, corresponding to the structure model (sites/occupancies). Elements not accounted for in the structural model should not be listed here. The given `_chemical_formula_weight` should match the formula given in `_chemical_formula_sum`.

If the cif also includes items that are derived from the chemical formula, such as density (_exptl_crystal_density_diffrn), scattering power (_exptl_crystal_F_000), or absorption coefficient (_exptl_absorpt_coefficient_mu) the values must be calculated from _chemical_formula_sum , and the _cell_formula_units_Z and the cell parameters.

If microprobe (or other) chemical analysis was performed, report the empirical formula (with esd's):

```
_chemical_formula_analytical        'K0.96(1) Al0.96(1) Si1.04(1) O4'
```

for minerals:

```
_chemical_name_mineral              'IMA approved mineral name'
_chemical_compound_source           'locality and/or collection reference'
                                    or 'synthesised at xx GPa and YY K'
```

4.3 Unit cell

All measured variable quantities must be accompanied by esd's.

```
_cell_formula_units_Z               3
_cell_length_a                      17.2760(19)
_cell_length_b                      35.957(5)
_cell_length_c                      7.2560(8)
_cell_angle_alpha                   90
_cell_angle_beta                    91.359(7)
_cell_angle_gamma                   90
_cell_volume                        4506.1(10)
```

if different from ambient conditions (units: K and kPa):

```
_cell_measurement_temperature       500(2)
_cell_measurement_pressure          1000000(50000)
```

Note that ambient pressure is 100 kPa, and it should be reported if the CIF also contains high-pressure data.

4.4 Experimental details

If the structure was determined/refined from non-ambient data (units: K and kPa):

```
_diffrn_ambient_temperature         1100(10)
_diffrn_ambient_pressure            2100000(50000)
```

Note that ambient pressure is 100 kPa, and it should be reported if the CIF also contains high-pressure data.

the used radiation must be defined by
_diffrn_radiation_probe
allowed data values are 'x-ray', 'neutron', 'electron' or 'gamma'

more details on the used radiation should be given using the following items:

```
_diffrn_radiation_type              'Mo K\a'
_diffrn_radiation_wavelength        0.71073
```

Furthermore, the minimum information must reveal the type of experiment (powder or single-crystal). This can be achieved as follows:

For powder diffraction experiments:

```
_pd_instr_geometry                  Bragg-Brentano
```

for single-crystal diffraction experiments:

```
_diffrn_measurement_device          'two-circle diffractometer'
_diffrn_measurement_device_type     'Stoe IPDS-2'
```

4.4.1 Simulated structures

Cif's for full-structure simulations are required and should contain all basic structural details. Instead of the information on the radiation and experimental setup, at least minimum details of the simulation should be given:

```
_computing_structure_solution       'software/method used'
_computing_structure_refinement     'software/method used'
_refine_special_details
; some basic details on the simulation method and parameters.
;
```

Furthermore, the simulation conditions should be given in

`_cell_measurement_temperature, _cell_measurement_pressure`

and

`_diffrn_ambient_temperature, _diffrn_ambient_pressure`

4.5 Symmetry information

```
_space_group_crystal_system         monoclinic
_symmetry_space_group_name_H-M      'C 1 m 1'
```

Please note, that `_symmetry_cell_setting` is deprecated and replaced by `_space_group_crystal_system`. The item `_space_group_name_H-M_alt` is also allowed.

To eliminate any ambiguity related to cell setting and origin the explicit declaration of the symmetry operators is required:

```
loop_
_symmetry_equiv_pos_as_xyz
x,y,z
x,-y,z
x+1/2,y+1/2,z
x+1/2,-y+1/2,z
```

`_space_group_symop_operation_xyz` is the new alternative to `_symmetry_equiv_pos_as_xyz`

4.6 Structural data

The fractional coordinates and occupancy factors are given in a loop, containing a minimum information of: `_atom_site_label`, `_atom_site_type_symbol`, `_atom_site_fract_x`, `_atom_site_fract_y`, `_atom_site_fract_z`, `_atom_site_adp_type`, `_atom_site_U_iso_or_equiv`, `_atom_site_occupancy`

All experimentally derived and/or refined parameters must be given with esd's! For more details and examples see the example files.

For simulated structures, displacement parameters are not usually determined. But the items `_atom_site_label` and/or `_atom_site_type_symbol`, plus `_atom_site_fract_x`, `_atom_site_fract_y`, `_atom_site_fract_z`, are required.

4.7 Data consistency

The data presented in the manuscript has to be consistent (strictly) with the data included in the cif file. Particularly, this applies to lattice parameters, atomic coordinates (names and labels), and ADP's, but also to all other values and derived values (distances, angles, BVS, etc.) Create and validate the cif first and then use available software to generate any Tables for the manuscript.

4.8 Empty data / comments

One of the advantages of cif files is that they are still human-readable. To optimise the readability, we recommend to remove all comment lines which are not considered to be useful to the reader (such as default comments produced by various software).

Please delete irrelevant data items introduced automatically by software (e.g. `_atom_sites_solution_hydrogens` in a structure with no H atoms).

Furthermore, please also remove all empty data items (usually denoted with ?), as they are useless. And remove as well empty `loop_` structures with no content.

4.9 Bibliographic information

Please do not include any bibliographic information in the CIF, as the needed items (to refer to the published article) will be added by the Am. Min. editorial office automatically. This includes all `_publ_*` and `_journal_*` items.

5. ADDITIONAL INFORMATION POLICY

Authors often decide to include more information into the cif file than the “minimal requirements”. We explicitly encourage authors to include as much information as possible; the more data that you include in the CIF, the more valuable the data is to other researchers, and the higher the likelihood that your results will be used and cited by others! Please make sure that all of the data is consistent with the manuscript (distances/angles, refinement details, etc.)

An example of a CIF that includes extended information, `maximal_cif_example.cif`, is available at http://www.minsocam.org/MSA/AmMin/CIF_info.html

5.1 Structure factors

If structure factors are submitted, they must be formatted using the cif syntax and to be included in the cif file to make sure they can easily be used by other authors.

6. FAQs, TRICKS AND TIPS

This section contains a number of suggestions for presenting data in CIF’s, and how to address problems that are identified in your CIF by the IUCr cifChecker [3]. It is based on the accumulated experience of the American Mineralogist Technical Editing Team, and the suggestions of authors such as you. New contributions are welcome!

6.1 Syntax problems

The IUCr cifChecker will often identify syntax problems in your CIF. These are often because of mismatched ‘;’ used to delimit multi-line text entries, or incorrect `loop_` instructions. Use a CIF-specific editor such as enCIFer [4] or publCIF [5] to look at your file and find the syntax error.

Lines in a CIF must be no longer than 80 characters. Because some software can get confused over what constitutes the length of a line, it is good practice to keep all lines well within the 80-character restriction. It is not necessary to pad lines out to a width of 80 characters with trailing spaces. Avoid the use of tab characters, which are sometimes transformed by e-mail into groups of up to eight space characters – something that can cause the line-length limits to be exceeded. Lines over 80 characters can be identified with the cif_fix program [6] or a CIF-specific editor such as enCIFer [4].

6.2 Invalid characters

To ensure portability and readability CIF’s must be text files containing only the printable ASCII characters. (these are defined as characters from the ASCII set with decimal integer values from 32 to 126, as well as horizontal and vertical tabs, line feed, form feed and carriage return characters). This includes the standard alphanumeric and punctuation characters on a US computer keyboard, but it excludes any accented letters or characters that may be available on international keyboards.

Do not open your CIF with an editor such as Word, as this may also add additional non-CIF content to the file, or illegal characters. Under Windows, either use a CIF-specific editor [4,5] or simpler editors such as Notepad or Wordpad.

If the IUCr cifChecker [3] reports invalid characters, they can be removed in several ways:

- In Unix: `sed -i 's/[^\x20-\x7e\x09\x0a\x0d]//g' filename.cif`
- Under Windows, use the cif_fix program [6].
- Using a CIF-specific editor such as enCIFer [4].

6.3 Multiple structures in one CIF

For ease of data handling American Mineralogist requires that all structures reported in a manuscript are provided in a single CIF. This is achieved by using 'data blocks'. In brief:

- All of the data for one structure refinement goes into one data block.
- Each data block must contain the complete information from one structure refinement. See section 4 for minimal requirements.
- Each data block must have a unique name, preferably meaningful.
- No data block with refinement results can be called `data_global`

The overall layout of the CIF will therefore look like this:

```
data_roompressure
_diffrn_ambient_temperature          300(3)
_diffrn_ambient_pressure             100(2)
_chemical_formula_sum                'Si O2'
_chemical_formula_weight             60.08
```

(etc – all of the information for this structure refinement at atmospheric pressure)

```
data_P1
_diffrn_ambient_temperature          300(3)
_diffrn_ambient_pressure             900000(2000)
_chemical_formula_sum                'Si O2'
_chemical_formula_weight             60.08
```

(etc – all of the information for this structure refinement at this pressure)

You should first create each data_block as a separate CIF and validate them separately (see section 2). To make one CIF, you can edit them together with any of the recommended editors. Or, you can simply concatenate the files:

- Under Windows:
  - Put all of the files in one folder. Make sure they all have the extension `.cif`
  - Open a command prompt box (i.e. Dos box) in the same folder
  - Issue the command '`copy *.cif  alldata.cif`' This will create a CIF with the name alldata.cif that contains all of the data blocks.
- In Unix, use the command: `cat *.cif >all.cif`

6.4 How to report multiple site occupancies: simple case

The key to doing this easily is to do the refinement with multiple site occupancies, so that the composition calculated from the site occupancies matches that given in the data item `_chemical_formula_sum`. Then your refinement will be physically-reasonable, your refinement program will report the multiple occupancies correctly in the CIF, without any problems for you!

But if you have to do it yourself, here is an example of how to arrange the loop with the atomic coordinates. The text after the # symbols are explanatory comments, and should not appear in a real CIF:

```
loop_
_atom_site_label                  # This is the name of the site
_atom_site_type_symbol            # This is the element symbol
_atom_site_fract_x
_atom_site_fract_y
_atom_site_fract_z
_atom_site_occupancy              # The occupancy of this element on this site
_atom_site_B_iso_or_equiv
_atom_site_adp_type
T1S Si 0.556(2) 0.1681(9) 0.2033(7) 0.5  1.06  Biso
T1A Al 0.556(2) 0.1681(9) 0.2033(7) 0.5  1.06  Biso
O1  O  0.123(7) 0.5000    0.431(1)  1.0  1.20  Biso
```

So, the example above describes a site called T1 that is 50% occupied by Al and 50% by Si. For fully occupied sites, the site occupancies must sum to 1.0. These are *not* the site occupation factors used in SHELX, which include the equipoint fraction.

You cannot use the label of just *T1* for `_atom_site_label` for both the Si and Al content of T1 because many programs (including encifer and checkcif) do not accept duplicate site labels. The construction suggested here, adding an extra identifier *A* and *S* after the traditional mineralogical site name T1 allows the traditional site name to be used, and to avoid these problems with chemistry-based programs. However, this is just a recommendation as there is no ‘standard’ way to handle this situation at this time; we welcome suggestions of alternative methods that also pass checkcif.

Note also that when this construction is used, the occupancy of fully-occupied sites, like the ‘O1’ in the example, must also be given explicitly.

Please make sure that lines for one site are next to each other and use exactly (!) the same coordinates.

If you list bond lengths and angles in the cif, make sure that you only list them for one of the T1 atoms, not both. Also eliminate any 'silly' bond lengths (such as T1A-T1S) and angles (e.g. T1A-O1-T1S). This can be done automatically in most refinement programs by setting limits to the bond lengths and angles calculation.

6.4.1 Multiple site occupancies in the manuscript

We strongly recommend that you use the site occupancies (and *not* the site occupation factors) when you are discussing and reporting the elements present on a site in the manuscript and its Tables. This avoids confusing readers. You can take the site occupancy values directly from the CIF. For fully occupied sites, the site occupancies must sum to 1.0. The reason for this recommendation is that site occupation factors can mislead readers into thinking that a fully-occupied site on a special position is in fact only partially occupied.

6.5 How to report multiple site occupancies: complex cases

In many mineral crystal structures, cation sites are occupied by a complex mixture of cations. A simple example would be a natural olivine in which the two M sites contain trace amounts of Ca, Mn, and Ni in addition to the major elements Mg and Fe. Even if the total site occupancy is constrained, it then becomes impossible to obtain a unique set of site occupancies for all elements simply by refinement to diffraction data, because the data only allow the total scattering power of each site to be constrained. Here are two possible approaches to this problem….

6.5.1 Apply external constraints

It may be possible from knowledge of the crystal chemistry, or from non-diffraction data, to assign some elements to one or more specific cation sites in the structure. The structure model can then be built with these constraints, and the multiple site occupancies reported as described in section 6.4. It may also be appropriate to restrain or constrain the refinement model to compositional data from a microprobe analysis. The constraints and restraints should then be described in the CIF in the section `_refine_special_details`. Then make sure that the composition calculated from the site occupancies matches that given in the data item `_chemical_formula_sum`. The full composition (if known from other methods) should be reported in `_chemical_formula_analytical`.

6.5.2 Group similar cations

If no external constraints are available, the refinement can be done in a physically-reasonable manner by grouping together cations with similar scattering power. For the example of olivine described above, the X-ray scattering factors of Ni, Mn and Fe are very similar. For trace amounts of Ni and Mn, no significant error is introduced into the other refined variables by including them with Fe in the structural model. Ca has a scattering power for X-rays intermediate between Mg and Fe. So, a refinement model in which only Fe and Mg are used then returns Fe occupancies that are really representative of the total Fe+Mn+Ni+1/2Ca. In this case, the composition calculated from the site occupancies reported in the data item `_chemical_formula_sum` will contain only Fe and Mg (in addition to Si and O), while the full composition (if known from other methods) should be reported in `_chemical_formula_analytical`.The explanation of the refinement model should be given in the CIF in the section `_refine_special_details`.

6.6 Un-located hydrogen atoms

We recommend that, if possible, a physically-reasonable model for hydrogen atom positions should be included whenever possible, because this ensures that the resulting crystal structure model is charge-balanced and reflects the true composition. Various methods exist for estimating and then constraining or restraining H-atom positions within a refinement if their positions cannot be freely and meaningfully refined to the available data. In all cases of restraint and constraint, the details should be reported in the CIF in the section `_refine_special_details`.

If the H-atoms cannot be included in the refinement model, even strongly constrained, or they cannot be reasonably modelled (e.g. for reasons of high disorder), the H should not be included in `_chemical_formula_sum`, but should be reported in `_chemical_formula_analytical`. An explanation should be provided in the CIF in the section `_refine_special_details`.