CIF Data Importer

Overview and instructions for installation and use

Also see www.rsc.org/Publishing/ReSourCe/AuthorGuidelines/AuthoringTools/CIFDataImporter/index.asp

CONTENTS

1. Overview of CIF Data Importer

The CIF data importer is a tool available in the RSC Word templates for authors to import crystallographic structure data from a CIF file as a formatted table or footnote.

Automating this task will save authors time and effort, and also reduce the opportunity for copying errors to be introduced.

It is currently only available for Word 2003/2004 operated on a PC. Problems may occur when running the tool on a Macintosh computer.

Please note that the CIF Data Importer is currently a Beta version for testing, and the RSC would welcome feedback by users that might improve it.

2. How to download and install the Word author templates

These are the steps that need to be performed before using the latest Microsoft Word author templates for the first time. We assume that the templates are being used on a PC (rather than a Macintosh) and the Microsoft Word version is 2003/2004.

1 Download the appropriate template zip file from the Microsoft Word Templates page of the RSC website to your computer and unzip it if necessary.

2 Decide which template to install – if you arre using a PC and wish to use the CIF Data Importer then choose a RSC_W2003_XXX_CIF_32.dot template (rather than a RSC_W2003_XXX_32.dot template, which are for those using Macintosh computers whch do not contain the CIF Data importer). XXX determines whether you are writing a communication, article or report.

3 Insert the template '.dot' file into your Word templates folder. If you are unsure where this is then open Word, look in the *Tools* menu, choose *Options* and view the *File Locations* tab to find the path to your User Templates directory. It is into here that you need to move the relevant template '.dot' file to.

4 Ensure that macros are enabled in Word - the macros that are embedded in the RSC template are there to help you and are not in any way dangerous. Go to the *Tools* menu, choose *Macro*, then *Security*... and ensure that it is set to *Medium* (under the *Security Level* tab). If not then click on that option to select it. You will need to exit and reopen Word for your changes to be applied.

5 To start writing a new article or communication or report with the guidance of the relevant template and RSC toolbar, open Word then, from the *File* menu and *New* window, select the relevant RSC template and open a new file (as with any other Word template).

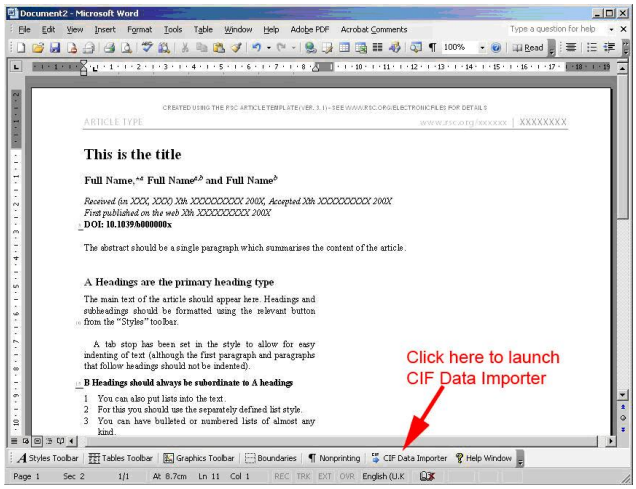
3. How to run the CIF Data Importer

Follow these steps to insert a table or footnote containing data imported from a CIF file, once the relevant Microsoft Word author templates have been installed and used to start writing a new document.

1 Check that your CIF file is formatted correctly. You may wish to use checking software from the CCDC http://www.ccdc.cam.ac.uk/free_services/encifer/

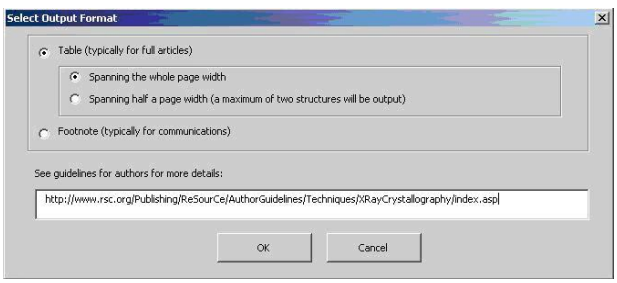
2 Position the cursor in the Word document at the point where the data is to be inserted.

3 Click the 'CIF Data Importer' button in the RSC toolbar at the bottom of the Word window.

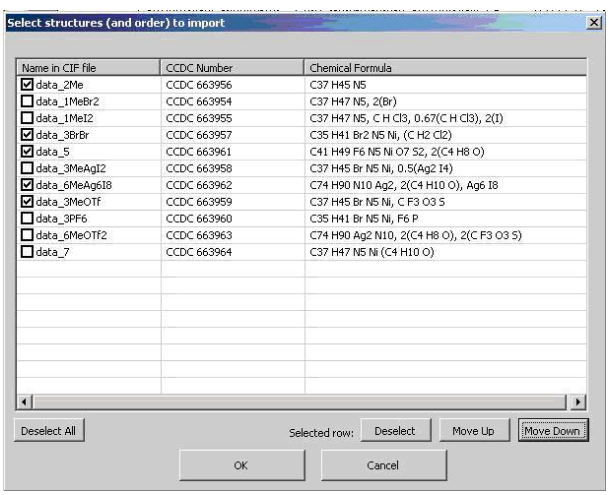


4 You will be presented with a dialog box to browse to the location of the CIF file(s) to import and select it for import. Note that while multiple .cif or .txt files can be selected for import, when submitting to the RSC it is preferred that the data for all structures is collated into a single .cif file.

5 Click the 'CIF Data Importer' button in the RSC toolbar at the bottom of the Word window.

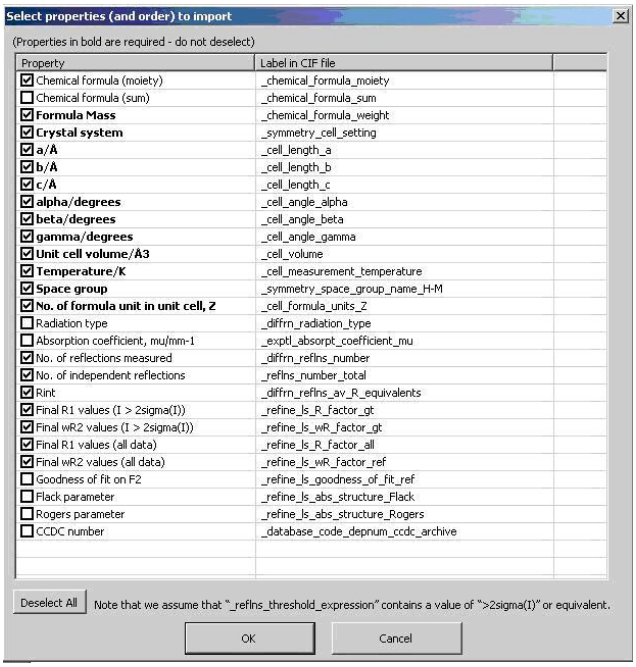


6 You will then be asked which structures in the CIF file you wish to import. Note that while a maximum of 20 structures can be imported that a table containing more than 10 structures starts to look cluttered. It is also possible to reorder them if their order in the CIF file is not the order that you wish to present them in.



7 If you wish the data to be displayed with an alternative name to the default (taken from the CIF file) then it is possible to type it in the following dialog box.

8 The final dialog box defines which properties from the CIF file you wish to import (and the CIF labels that correspond to them). Note that properties in bold are required and should not be removed from the output.



- Once your table or footnote has been created you should check it in Microsoft Word. You may wish to reformat some parts of it e.g. add subscripts to the space group names to make the screw axes clearer.

4. Information about CIF format required

Multiple CIF files can be selected for import if the structures to be imported are not in a single file (although for submission it would be preferred if these could be compiled into a single CIF file). Files with extension .cif and .txt can be selected for import.

Different crystal structures start with a line which starts with '*data_*' (with the exception of the '*data_global*' and '*data_publication_text*' lines) and end with '*#==END*' (although if this label is missing the import will still work). The suffix following '*data_*' is taken to be the default name for the structure (e.g. *data_NameInCIFfile)*, but an alternative name can be input if desired.

The following properties are required for the crystallographic data imported from the CIF file to be valid (the label from the CIF file is given in brackets):

- Chemical formula (*_chemical_formula_moiety* or *_chemical_formula_sum*)
- Formula Mass (*_chemical_formula_weight*)
- Crystal system (*_symmetry_cell_setting*)
- *a*/Å (*_cell_length_a*)
- *b*/Å (*_cell_length_b*)
- *c*/Å (*_cell_length_c*)
- *α*/° (*_cell_angle_alpha*)

- β/° (*_cell_angle_beta*)
- γ/° (*_cell_angle_gamma*)
- Unit cell volume/Å (*_cell_volume*)
- Temperature/K (*_cell_measurement_temperature*)
- Space group (*_symmetry_space_group_name_H-M* or *_space_group_name_H-M_alt*)
- No. of formula units per unit cell, *Z* (*_cell_formula_units_Z*)

At least one of the following properties are required for the crystallographic data imported from the CIF file to be valid:

- No. of reflections measured (*_diffrn_reflns_number*)
- No. of independent reflections (*_reflns_number_total*)
- R_{int} (*_diffrn_reflns_av_R_equivalents*)

At least one of the following properties are required for the crystallographic data imported from the CIF file to be valid (note that we assume that '*_reflns_threshold_expression*' contains a value of '>2sigma(I)' or equivalent):

- Final R_1 values ($I > 2\sigma(I)$) (*_refine_ls_R_factor_gt*)
- Final $wR(F^2)$ values ($I > 2\sigma(I)$) (*_refine_ls_wR_factor_gt*)
- Final R_1 values (all data) (*_refine_ls_R_factor_all*)
- Final $wR(F^2)$ values (all data) (*_refine_ls_wR_factor_ref*)

The following properties can optionally be added:

- Absorption coefficient, μ/mm (*_exptl_absorpt_coefficient_mu*)
- Radiation type (*_diffrn_radiation_type*)
- Goodness of fit on F^2 (*_refine_ls_goodness_of_fit_ref*)
- Rogers parameter (*_refine_ls_abs_structure_Rogers*)
- CCDC number (*_database_code_depnum_ccdc_archive*)

In general, the appropriate property values are simply extracted from the CIF file and inserted into the article with no processing. However, the following exceptions apply:

- Chemical formula (from the *_chemical_formula_moiety* or *_chemical_formula_sum*) has:
  - spaces removed from it
  - commas replaced by the special character •
  - numbers which don't immediately preceed an open bracket (or [character are turned into subscripts
- Radiation type: the string '\a' is replaced by 'α'
- Space groups: non-numeric characters are italicised, spaces are deleted, and the following substitutions applied:
  - -1 is converted into $\bar{1}$
  - -3 is converted into $\bar{3}$
  - -4 is converted into $\bar{4}$
  - -6 is converted into $\bar{6}$

5. Further information

About RSC Microsoft Word templates:
http://www.rsc.org/Publishing/Journals/guidelines/AuthorGuidelines/AuthoringTools/Templates/word.asp
About the CCDC checking tool encipher:
http://www.ccdc.cam.ac.uk/free_services/encifer/

6. Examples of formatted data

An example of data imported and formatted as a footnote is:
[a]Crystal data for 1a: $C_{12}H_{14}Cl_2N_2$, M = 257.15, triclinic, a = 6.5705(11) Å, b = 7.6756(14) Å, c = 12.6098(18) Å, α = 85.165(13)°, β = 76.785(13)°, γ = 73.868(16)°, V = 594.58(17) Å^3, T = 120(2) K, space group $P\bar{1}$, Z = 2, μ(MoKα) = 0.519 mm^{-1}, 14508 reflections measured, 3452 independent reflections (R_{int} = 0.0107). The final R_1 values were 0.0218 ($I > 2\sigma(I)$). The final $wR(F^2)$ values were 0.0708 ($I > 2\sigma(I)$). The final R_1 values were 0.0243 (all data). The final $wR(F^2)$ values were 0.0734 (all data). The goodness of fit on F^2 was 1.115. CCDC number CCDC 618057. Crystal data for 1b: $C_{12}H_{14}Cl_2N_2$, M = 257.15, triclinic, a = 4.3029(9) Å, b = 5.7033(9) Å, c = 12.6098(18) Å, α = 100.859(13)°, β = 95.703(13)°, γ = 99.229(11)°, V = 297.29(9) Å^3, T = 120(2) K, space group $P\bar{1}$, Z = 1, μ(MoKα) = 0.519 mm^{-1}, 7056 reflections measured, 1721 independent reflections (R_{int} = 0.0097). The final R_1 values were 0.0322 ($I > 2\sigma(I)$). The final $wR(F^2)$ values were 0.0906 ($I > 2\sigma(I)$). The final R_1 values were 0.0328 (all data). The final $wR(F^2)$ values were 0.0912 (all data). The goodness of fit on F^2 was 1.068. CCDC number CCDC 618059. Crystal data for 2: $C_{12}H_{14}Cl_2N_2$, M = 257.15, orthorhombic, a = 27.4793(11) Å, b = 6.0280(3) Å, c = 7.3190(3) Å, α = 90.00°, β = 90.00°, γ = 90.00°, V = 1212.36(9) Å^3, T = 120(2) K, space group *Pbcn*, Z = 4, μ(MoKα) = 0.509 mm^{-1}, 37294 reflections measured, 1776 independent reflections (R_{int} = 0.0162). The final R_1 values were 0.0249 ($I > 2\sigma(I)$). The final $wR(F^2)$ values were 0.0785 ($I > 2\sigma(I)$). The final R_1 values were 0.0294 (all data). The final $wR(F^2)$ values were 0.0807 (all data). The goodness of fit on F^2 was 1.059. CCDC number CCDC 618061.

An example of data imported and formatted as a table is:

Table XX Caption

Compound reference	2Me	3BrBr	5	6MeAg6I8	3MeOTf
Chemical formula	$C_{37}H_{45}N_5$	$C_{35}H_{41}Br_2N_5Ni•(CH_2Cl_2)$	$C_{41}H_{49}F_6N_5NiO_7S_2•2(C_4H_8O)$	$C_{74}H_{90}N_{10}Ag_2•2(C_4H_{10}O)•Ag_6I_8$	$C_{37}H_{45}BrN_5Ni•CF_3O_3S$
Formula Mass	559.78	835.18	1104.89	3145.96	847.47
Crystal system	Monoclinic	Triclinic	Monoclinic	Monoclinic	Triclinic
a/Å	15.058(3)	12.7728(5)	22.7194(19)	13.6353(4)	8.6147(3)
b/Å	24.920(6)	13.6700(5)	25.510(3)	22.1211(6)	14.2268(7)
c/Å	10.967(3)	25.2292(10)	14.1716(13)	17.0640(5)	16.2224(8)
α/°	90.00	87.286(2)	90.00	90.00	101.802(2)
β/°	91.360(3)	75.428(2)	125.520(4)	98.0900(10)	98.787(3)
γ/°	90.00	71.776(2)	90.00	90.00	96.327(3)
Unit cell volume/Å^3	4114.2(17)	4047.2(3)	6685.2(11)	5095.8(3)	1902.67(15)
Temperature/K	120(2)	120(2)	120(2)	120(2)	120(2)
Space group	*P*21/*c*	$P\bar{1}$	*C*2/*c*	*P*21/*n*	$P\bar{1}$
No. of formula units per unit cell, *Z*	4	4	4	2	2
Radiation type	Synchrotron	MoKα	MoKα	MoKα	MoKα
Absorption coefficient, μ/mm^{-1}	0.054	2.616	0.415	3.972	1.673
No. of reflections measured	22457	69924	28442	78859	27224
No. of independent reflections	5879	18589	7590	11729	6686
R_{int}	0.0579	0.1227	0.0715	0.0516	0.1479
Final R_1 values ($I > 2\sigma(I)$)	0.0700	0.0900	0.0869	0.0560	0.1050
Final $wR(F^2)$ values ($I > 2\sigma(I)$)	0.1937	0.1668	0.1959	0.1026	0.1744
Final R_1 values (all data)	0.1048	0.1837	0.1390	0.0753	0.1853
Final $wR(F^2)$ values (all data)	0.2125	0.1947	0.2181	0.1121	0.2085
Goodness of fit on F^2	0.976	1.008	1.056	1.166	1.126
Flack parameter					
Rogers parameter					
CCDC number	CCDC 663956	CCDC 663957	CCDC 663961	CCDC 663962	CCDC 663959