

CIF Applications. III. *CYCLOPS*: for Validating CIF Data Names*

BY SYDNEY R. HALL

Crystallography Centre, University of Western Australia, Nedlands 6009, Australia

(Received 12 December 1992; accepted 15 January 1993)

Abstract

The program *CYCLOPS* is used, in conjunction with a CIF Dictionary, to validate data names in an ASCII file. This file may contain CIF or non-CIF data, text documents or a program source. *CYCLOPS* is written in Fortran77 and has been implemented on a wide range of computers. It is available as public-domain software.

Introduction

Each data item stored in a Crystallographic Information File (CIF; Hall, Allen & Brown, 1991) is identified by a unique data name. The correct spelling of this data name is essential if the associated data value is to be accessed. In fact, the basic function of all CIF applications is to parse ASCII data searching for recognizable data names. Incorrect construction or spelling will make data inaccessible. The role of *CYCLOPS* is to validate data names against a CIF dictionary. In doing so, *CYCLOPS* highlights the occurrence (frequency and location) of all data names present in a file.

CYCLOPS has been used extensively in a number of software developments to ensure that data names conform with the current CIF definitions contained in a dictionary file constructed in the CIF Dictionary Data Language format (DDL; Cook, 1991). The CIF Core Dictionary file *cifdic.C91* is in this format.

Algorithm

The following procedure is used by *CYCLOPS* to check data names.

1. Read the dictionary file and store all data names in a sequential list. The dictionary file is opened as *STARDIC*.
2. Read the ASCII file to be checked. This is opened as the *stdin* file. For operating systems that do not support *stdin*, the file is opened as *STARTEXT*. Here are the procedure and criteria used in the data-name search.

(a) The file is parsed line by line for data names. These are recognized as any character string bounded by blanks and starting with an underline ‘_’ character.

* This paper is one of a series of papers on CIF applications. Offprints are available from The Technical Editor, 5 Abbey Square, Chester CH1 2HU, England. See text of paper for availability of program(s) by email.

Table 1. Sample output at the start of a *STARCHECK* file

<i>CYCLOPS</i> check list.	
Dictionary data names = 421.	
New data names in text = 13.	
CIF Dictionary (Core 1991)	Line numbers
_atom_site_aniso_label	1
_atom_site_aniso_type_symbol	2
_atom_site_aniso_u_11	3
_atom_site_aniso_u_12	
_atom_site_aniso_u_13	
_atom_site_aniso_u_22	
_atom_site_aniso_u_23	
_atom_site_aniso_u_33	
_atom_site_attached_hydrogens	4
_atom_site_calc_attached_atom	5
_atom_site_calc_flag	6
_atom_site_cartn_x	7
_atom_site_cartn_y	
_atom_site_cartn_z	
_atom_site_chemical_conn_number	8
_atom_site_constraints	9
_atom_site_description	10
_atom_site_disorder_group	11
_atom_site_fract_x	13
_atom_site_fract_y	14
_atom_site_fract_z	15
_atom_site_label	12
_atom_site_label_component_0	16
_atom_site_label_component_1	
_atom_site_label_component_2	
_atom_site_label_component_3	
_atom_site_label_component_4	
_atom_site_label_component_5	
_atom_site_label_component_6	
_atom_site_occupancy	17
_atom_site_refinement_flags	18
_atom_site_restraints	19
_atom_sites_solution_primary	20
_atom_sites_solution_secondary	
_atom_sites_solution_hydrogens	
_atom_site_symmetry_multiplicity	21
_atom_site_thermal_displace_type	22
_atom_site_type_symbol	23
_atom_site_u_iso_or_equiv	24
_atom_site_wyckoff_symbol	25
_atom_sites_cartn_tran_matrix_11	26
_atom_sites_cartn_tran_matrix_12	
_atom_sites_cartn_tran_matrix_13	
_atom_sites_cartn_tran_matrix_21	
_atom_sites_cartn_tran_matrix_22	
_atom_sites_cartn_tran_matrix_23	
_atom_sites_cartn_tran_matrix_31	
_atom_sites_cartn_tran_matrix_32	
_atom_sites_cartn_tran_matrix_33	
_atom_sites_cartn_transform_axes	27
_atom_type_analytical_mass_#	28

Table 2. Sample output at the end of a STARCHEK file

	Line numbers
CIF Dictionary (Core 1991)	
_reflns_scale_meas_f	249
_reflns_scale_meas_f_squared	
_reflns_scale_meas_intensity	
_reflns_special_details	250
_symmetry_cell_setting	251
_symmetry_equiv_pos_as_xyz	252
_symmetry_int_tables_number	253
_symmetry_space_group_name_hall	254
_symmetry_space_group_name_h-m	255
Data names NOT in Dictionary	
_cell_length_a_pm	42
_cell_length_a_nm	43
_cell_length_b_pm	45
_cell_length_b_nm	46
_cell_length_c_pm	48
_cell_length_c_nm	49
_cell_volume_pm	64
_cell_volume_nm	65
_chemical_melting_point_c	82
_diffrn_ambient_temperature_c	95
_geom_bond_distance_pm	174
_geom_bond_distance_nm	175
_journal_page	188

(b) Embedded data names are also recognized provided the underline is preceded by any of the following characters: , . : ([{ < / | " ' and the string is trailed by any of the following characters: , . ? ! ; :)] } > / - = | " ' .

(c) All alphabetic characters in the data names are subsumed to lower case.

(d) Data names are checked against the stored dictionary list. If a match is found, the line number of the ASCII file is stored. Up to 100 line numbers can be stored for each data name.

(e) If the data name is not present in the sequential list, it is added to the end of the list with its line number.

3. List the number of data names extracted from the dictionary file along with the number of data names that did not match this list.

4. Output the contents of the sequential data-name list with the line numbers as the file STARCHEK. Two extracts from a STARCHEK file are given in Tables 1 and 2.

Distribution

CYCLOPS is distributed as the file *cyclops* containing the Fortran source, the common file and a test script using the *cifdic.C91* dictionary file. The files *cyclops* and *cifdic.C91* may be obtained free of charge in several different ways. The simplest and fastest approach is to use anonymous FTP to *get* the file from the directory *cif* on the host 130.95.232.12. Alternatively, send an email containing the line `send cyclops cifdic.C91` to either `sendcif@crystal.uwa.edu.au` or `sendcif@iucr.ac.uk`. As a last resort, airmail a floppy disk to the author stating the mode of copy required.

References

- COOK, A. P. F. (1991). *Report on Implementing SMD in STAR: Dictionary Definition Language*. ORAC Ltd, Leeds, England.
- HALL, S. R., ALLEN, F. H. & BROWN, I. D. (1991). *Acta Cryst. A47*, 655–685.