

International Union of Crystallography

Electronic Data Entry Form for *World Directory of Crystallographers*, Ninth Edition

Abstract

A data entry form allowing biographical data to be entered in the *World Directory of Crystallographers* is described. The form is a STAR file allowing automated typesetting of the Directory and the establishment of an electronic database of crystallographers' names, addresses and interests.

Introduction

The International Union of Crystallography has since 1957 been responsible for producing eight editions of an international directory of crystallographers. The projected ninth edition of this work will be established as an electronic database, allowing constant updating of changing information. The database will be established from electronic files supplied by the national Subeditors who are responsible for their own country's entries. The master input files will be in the STAR format described by Hall [(1991). *J. Chem. Inf. Comput. Sci.* **31**, 326–333]. This file format is already used by the IUCr as the basis of the Crystallographic Information File (CIF) used by authors to transmit structural papers to the journal *Acta Crystallographica*. Software similar to that used to prepare journal articles from CIF input will be used to publish a typeset edition of the *World Directory*.

Individuals may transmit their own entries to their national Subeditor for inclusion in the *World Directory* using the template electronic file described in this article. Copies of the file may be obtained by network email from the Technical Editor's office by sending a message to the address sendcif@iucr.ac.uk (or, within the UK, sendcif@uk.ac.iucr) containing the line `send form.wdc`. (Please note that all mail sent to this address is handled electronically by an information server program, and so must conform to the syntax indicated.)

A list of national Subeditors may also be retrieved by sending the request `send subedit.lst` to the `sendcif` address.

Completion of data form

The electronic form distributed from the Technical Editor's Office is shown in the Appendix. Any text to the right of a hash character is treated by STAR file processing software as a comment and ignored. This allows the file to include comments and an example describing how it should be completed. Apart from the comments, the file includes a data block declaration (`data_wdc9`: the 'wdc9' is an arbitrary choice), a list of `datanames` to follow in a 'loop_' structure, and the data itself. The query marks in the latter part of the file serve two purposes – they indicate where the contributor's information should be placed, but they also act as valid (but information-free) data items, thus maintaining the integrity of the file.

Descriptions of the fields are given below in a Dictionary of allowed data names.

General guidelines

The structure of data stored in a STAR file is determined by the presence of data names indicating the nature of the data to follow (the 'S' in the acronym means 'self-defining'). In the data entry form file, a set of twelve data fields is declared within a loop structure. It is therefore essential for the proper parsing of this file that each entry contains twelve

data fields. In the blank template, twelve dummy fields are supplied as query marks (fields of type *text* are indicated with the appropriate delimiting semicolons). These query marks should be replaced with the contributor's personal details. If a particular category is not appropriate, the matching query mark should be left intact, and not deleted.

A field of type *char* is a continuous string of no more than 80 printable ASCII characters. It must completely fit on one line of text. If it represents a phrase, embedded blanks may be replaced by underscore characters, or matching single or double quote marks may bracket the phrase. For example, the two component family name 'Van Dyk' may be entered in the `_name_family` field as `Van_Dyk`, `'Van Dyk'` or `"Van Dyk"`. If the expression to be entered contains quote marks (as apostrophes or components of a code for an accented letter), quote marks of a different type should be used, or the conversion of blank spaces to underscores may also be performed. Thus, enter the name O'Neill as `"O'Neill"` and not `'O'Neill'`.

Longer text, extending over more than one line, may be entered in fields of type *text* which are bounded by a semicolon as the leftmost character of the first and trailing lines. The first line may contain the start of the text, or may be blank, according to taste, but the last line contains only the closing semicolon. Thus, the address

```
; Case 115
  75252 Paris Cedex 05
  France
```

; is entered correctly, as is the example

```
;
Case 115
75252 Paris Cedex 05
France
```

; but a final line saying `France` ; would be wrong, as would a last line with a space or tab character before the semicolon.

In the examples of text fields given above, the alignment of lines of text is for purely aesthetic reasons – lines need not be offset from the left margin, and can be given in free form, with the proviso that certain individual fields must be formatted according to the rules given in the relevant Dictionary entry.

Accented or other special characters may be indicated in any field using the coding scheme given in the Dictionary under the `_name_family` entry.

General Editor

The General Editor for the *World Directory* project will be pleased to answer any questions and give any assistance required. Crystallographers who do not have access to a national Subeditor may also send their data forms direct to him. The address of the General editor is:

Y. Epelboin
LMCP, Universités P. M. Curie et Paris VII,
URA 009, CNRS,
Case 115, 75252 Paris CEDEX 05,
France
[e-mail: epelboin@lmcp.jussieu.fr]

Dictionary of data names for the *World Directory*

Version: STAR Dictionary (WDC 1992)

_name_family (char)

Family name. The name used for primary identification. Multiple components should be hyphenated or capitalised according to local custom. If quote delimiters are used, they should not conflict with embedded diacritical marks. Accents and special alphabetic characters should be indicated as follows:

\` acute	\" umlaut	\= overbar
\^ grave	\~ tilde	\. overdot
\~ circumflex	\; hook (ogonek)	\< hacek ("vee")
\, cedilla	\> long umlaut	\(breve
\%a a-ring	\?i dotless i	\&s German "ss"
\/o o-slash	\/l barred l	\/d barred d

Dynastic components ('Jr', 'III') may also be given.

Example(s): Epelboin, 'Van Dyk', von_Graben, "O'Neill"

_name_other (char)

Given names or initials. Title may precede the name, and should be shortened with a trailing point if it is an abbreviation (e.g. 'Prof.' for 'Professor') and without a trailing point if it is a contraction ('Dr' for 'Doctor').

Example(s): 'Prof. Yves', 'Dr Richard J.', 'Mr Zhong-he'

_address_institution (text)

Institution where resident. This will usually include department or section. Addresses spanning several lines should not have trailing commas or full points for punctuation. This field is intended to characterize the institution to which you belong.

Example(s):

```
; Lab. Min\`eralogie--Cristallographie
Universit\`es P. M. Curie et Paris VII
U. A. 009 CNRS
;
```

_address_postal (text)

Full postal address for correspondence. May differ from _address_institution and may contain duplication of information from _address_institution, if appropriate. Addresses spanning several lines should not have trailing commas or full points for punctuation. This field must contain all information necessary for postal delivery. Country must be given as the only information on the last line.

Example(s):

```
; Universit\`e P.M. Curie
Case 115
75252 Paris Cedex 05
France
;
```

_phone (char)

Telephone number. Should be given in the form `<country code><area code><local number>[x<extension>]` where parentheses surround the area or regional code, and an extension number, if given, is prefixed by an 'x'.

Example(s): 33(1)44275211, 44(244)342878x27

_fax (char)

Fax number. Should be given in the form `<country code><area code><local number>` where parentheses surround the area or regional code.

Example(s): 33(1)44273785

_telex (char)

Telex number.

Example(s): '669755 OFFICE G attn UNICRYSTAL'

_email (char)

Email address. Should be given all in lower case unless the relevant network addressing software is known to be case-sensitive. Domain-type addresses should be given as

`<user>@<machine>.<site>.<net>.<country>`

(i.e. in little-endian order), but sites which have addresses on different networks may use alternative ordering where there is danger of ambiguity. Bitnet addresses should contain the routing suffix '.bitnet'. Addresses on networks other than Internet or Bitnet should give routing information *via* known Internet/Bitnet gateways, where possible.

Example(s): epelboin@lmcp.jussieu.fr, teeter@bcchem.bitnet, 'bm@iucr.ac.uk (Internet) bm@uk.ac.iucr (JANET)', xray%neosoft.ineos.free@suearn2.earn, ...!uucpl!uucp2!uucp3!addressee

_degree_highest (char)

Highest degree. Given in abbreviated form without periods. Avoid titles which are not well known outside your own country such as DiplMet, StudLic, IngDoct, CSc. Try to find an equivalent internationally known.

Example(s): DSc, PhD, DPhil, MS, MSc, Cand

_degree_institution (char)

Institution awarding degree. Use the abbreviations U. for University, Coll. for College, Inst. for Institute or Institution.

Example(s): 'Paris VI U.', "Tech. U. \\\L'od'z", 'Imperial Coll. London'

_degree_date (char)

Year in which highest degree was conferred.

_degree_field (char)

Degree field.

Example(s): Physics, 'Chemical crystallography'

_birth_date (char)

Year of birth.

_current_position (char)

Current position. Avoid titles peculiar to your institution or country, such as Akad. Oberrat, Head, Senior registrar, Chercheur. Describe your position in English and do not use abbreviations.

Example(s): Lecturer, Associate_professor, 'Post-Doctoral Research Assistant'

_interests_key_words (text)

Current interests by keywords. The entry should contain no more than four terms from the list of allowed keywords published in *Acta Cryst.* (1992), A48, 949-954. Compound terms may be generated by affixing keywords from the 'attributes' list to keywords from the 'methods' or 'compounds' lists. Words in a compound term should be separated by underscores; space or tab characters should separate terms.

Example(s):

```
; computing X-ray_topography silicon_compounds
;
```

_interests_other*(text)*

Interests not precisely covered by keywords in previous field. There is no limit on the number or nature of phrases that may be given. Words in a compound term should be separated by underscores; space or tab characters should separate terms. *Warning:* no database searches will be performed on this field.

Example(s):

```
; real-time_online_scattering_experiments  
computerised_typesetting  
;
```

_iucr_responsibilities*(text)*

Position(s) of responsibility within the International Union of Crystallography. Responsibility in Commissions and in National Unions may be indicated here.

Example(s):

```
; Chairman, Commission on Crystallographic Computing  
;
```

_permission_to_distribute*(char)*

Authorization for the International Union of Crystallography to distribute information in this entry.

Permitted responses are **yes** **no**. Where no value is given, the assumed value is 'yes'.

APPENDIX: Listing of the data entry form

```

#####
### Data Entry Form for World Directory of Crystallographers ###
###                               (Ninth Edition)                               ###
#####
#
# This is an electronic "form" for entries into the 9th Edition
# of the IUCr World Directory of Crystallographers. This file
# will be used for automatic typesetting of a printed version
# of the World Directory, and as a basis for electronic access
# to this information.
#
# Please fill in the details as indicated below and send to the
# appropriate Sub-editor. This data is entered in the format of
# a STAR File. If you require further details on this type of
# file structure refer to the papers (Hall (1991) J Chem Inf
# Comp Sci 31,326-333) or (Hall, Allen & Brown (1991) Acta Cryst
# A47,655-685.)
#
#-----
# The following items are requested for each entry. These are
# given here as a STAR data 'name' and a brief description. Note
# that a 'char' data item is a string of less than 80 characters
# on the same line -- if this string contains imbedded blanks
# then it must be bounded by single or double quotes. If the data
# item is of type 'text' it must be bounded by semicolons as the
# first character of a line. An example entry is provided. Note
# that missing data items are flagged with a single question mark.
#-----

data_wdc9 # this is a STAR data block identifier
loop_     # this is STAR command to repeat the following data items
_name_family #char Name used for primary identification
_name_other  #char Other name(s). Title is optional.
_address_institution #text Primary institution address.
_address_postal #text Postal address
_phone        #char <country><city><local>x<extension>
_fax         #char Same format as phone number.
_telex       #char Complete international telex number.
_email       #char <namecode>@<host>.<net>.<country>
_degree_highest #char Highest degree received. Initials.
_degree_institution #char Institution where degree received.
_degree_date   #char Year that the degree was awarded.
_degree_field  #char Field of study for the degree.
_birth_date   #char Year of birth.
_current_position #char Position in the above institution.
_interests_key_words #text Key words* describing interests.
_interests_other  #text Interests other than key words*.
_iucr_responsibilities #text Current positions held in the IUCr.
_permission_to_distribute #char 'yes' or 'no' to electronic distr.

#                               * See (Acta Cryst (1992) A48,949-954).
#
#----- EXAMPLE ENTRY -----
# 'Epelboin'
# 'Prof Yves'
# ;Lab. Min\`eralogie-Cristallographie
# Universit\`es P. M. Curie et Paris VII
# U.A. 009 CNRS
# ;
# ;Case 115
# 75252 Paris Cedex 05
# France
# ;

# 33(1)44275211
# 33(1)44273785
# ?
# 'epelboin@lmcp.jussieu.fr'
# DSc
# 'Paris VI U'
# 1974
# 'Physics'
# 1944
# 'Group leader'
# ;computing X-ray_topography silicon_compounds
# ;
# ;Scientific_visualization
# ;
# ;Editor of ninth edition of World Directory of Crystallographers
# ;
# yes
#
#----- PLEASE FILL IN THE FIELDS BELOW FOR YOUR OWN ENTRY -----
# family name
# ?
# other name(s)
# ?
# institution
# ; ?
# ;
# postal address
# ; ?
# ;
# phone
# ?
# fax
# ?
# telex
# ?
# email
# ?
# highest degree
# ?
# degree institution
# ?
# degree date
# ?
# degree field
# ?
# degree field
# ?
# year of birth
# ?
# current position
# ?
# interests (by keyword)
# ; ?
# ;
# other interests
# ; ?
# ;
# IUCr responsibilities
# ; ?
# ;
# permission to distribute
# ?
#-----

```

This is an example of the data entry form (slightly modified for typographical representation). It is a complete STAR file, defined by a datablock name identifier, loop declaration and list of data items. All data items currently have dummy values (indicated by query marks), *but their presence as placeholders is nevertheless essential to ensure the logical integrity of the file structure*. The author should *replace* each query mark with the appropriate information. For text fields, where information may extend over more than one line, it is *essential* that the delimiting semicolons are retained at the leftmost positions of the initial and concluding lines. All text on a line to the right of a hash character (#) is treated as a comment, and ignored by file processing software. This allows an explanatory text and example to be included in the file.