# Accessing the WDC9 Database

**Y. Epelboin**
**WDC9 General Editor, L.M.C.P. Universite P.M. Curie, case 115, 75252**
**Paris Cedex 05, France [epelboin@lmcp.jussieu.fr]**

**B. Mc Mahon**
**Officer for Research & Development, IUCr, 5 Abbey Square,**
**Chester CH1 CHU, England [bm@iucr.ac.uk]**

**S.R. Hall**
**Crystallography Center, University of Western Australia,**
**Nedlands 6009, Australia [syd@crystal.uwa.edu.au]**

## *Synopsis*

A database on scientists with crystallographic interests has been established and located at the IUCr Office in Chester (UK). Methods for remotely accessing this database are described.

## *Introduction*

A database on scientists with crystallographic interests has been constructed (Epelboin, 1992a). National sub-editors have been assigned by the IUCr to collect information on addresses and research interests of scientists in each country. This data has been collated by Regional Editors for America, Asia and Europe to form the WDC9 database. The format of the database complies with that of the STAR File (Hall, 1991) which is the accepted IUCr standard for exchanging and archiving data. A hardcopy version of the WDC database will be printed every two years.

The contents of the WDC9 database have been described by McMahon (1993). Appendix A lists the data items stored in the database. The research interests of scientists are listed in terms of standard key-words (Epelboin, 1992b). This permits searches on particular research topics.

Access to the WDC database requires a registered security code, and a connection to an international computer network. The several different access approaches which are available for searching the WDC data base are described below. The appropriateness of these approaches depends on the application and the network connection. Those with minimum facilities can use e-mail access, while those with an Internet connection may use several modes of interactive access.

## *E-mail and telnet access modes*

### E-mail mode

E-mail queries to the WDC database are sent to:
**wdc-search@diamond.iucr.ac.uk**

The WDC query language used for e-mail access is discussed below, and details are given in Appendix C.

**Telnet mode**

Interactive telnet access to the WDC database is initiated with:

**telnet diamond.iucr.ac.uk**

The telnet session proceeds as follows:

(i) Enter the login sequence **wdc** when prompted.

(ii) Do not enter any password when prompted, just a carriage return.

(iii) Enter your query as a sequence of commands. Close the query sequence with a control-D. A detailed description of the WDC query language is given in Appendix C. Here are some additional hints.

- The "input file" prompt refers to a pre-prepared query sequence stored on a separate file. This approach is useful for multiple or repeated queries. The usual response is "no".
- The query input by the user will be displayed on the screen for checking. Confirm if it is correct.
- The search results of the query will also appear on the screen, but can be redirected to your email address.

**WDC query language**

E-mail and telnet searches use a common query language. This was developed for the WDC access software by Mark Favas and Syd Hall, and is a subset of the general *Star_Base* query language (Spadaccini & Hall, 1994).

A quick introduction to WDC query language will be provided via several annotated examples. These examples show how a "Dr Maxwell" would retrieve information from the WDC database using e-mail or telnet.

Example 1

Dr Maxwell needs information about a person with the family name "van Bloggs". Here are the commands used to do this search.

(i) The sequence "user", followed by the security code and name of the user must be used to open a search session. The combination of the code and name must match with an entry in the WDC database. The security code is supplied to registered users by the IUCr office.

**user ukxxxfd0876 Maxwell**

Queries are case insensitive. Each security code has a pre-assigned access limit for the number of successful matches (i.e. "hits") per month. For most users this limit is 150.

(ii) The following sequence is used to request information on all persons in the world named "van Bloggs". Note that some knowledge of the WDC data names will be needed to request a data item (such as _name_family). These are listed in Appendix A.

**if _name_family == vanbloggs**

Again, the search string is case insensitive, and all blanks and special accents are ignored. The "double equals" operator specifies an "exact match" with "vanbloggs". If one was uncertain of the exact spelling, the use of "~= vanblog" will search for partial matches. All search operators are listed in Appendix C.

(iii) The search is started with the specific command

**search**

This will result in all information about persons with the family name "vanbloggs" to be returned as e-mail output or to the screen. Multiple searches are possible within the same e-mail or telnet session.

Example 2

In this example Dr Maxwell searches for scientists with similar research interests in a specific country.

(i) The allocated number of search matches per month is fixed for each user. To prevent excessive hits in a search, the hit limit is set to 10.

**user ukxxxfd0876 Maxwell**
**limit 10**

(ii) Because Dr Maxwell only wants to know about scientists working in France, the search is constrained by entering:

**location fr**

The abbreviated code for each country is listed in Appendix B. Regional codes for America, Asia and Europe may also be used.

(iii) The research interests of each scientist are stored as the two data items **_interests_keyword** and **_interests_other**. In this example the specific topic of interest is "incommensurate structures". Because key-word combinations exist in the WDC, the best form of a search query is:

**if _interest* ~= incommensur**

The wild card character "*" is used to signal that both "interests" data items should be searched for the string "incommensur".

(iv) Dr Maxwell only wants the names, institution and interests of people satisfying the above. This is requested with the following command.

**want _name* _address_institution _interest***
**search**

The "want" command is only used when specific items are needed. The default request is for all data (i.e. _*), as was seen in the first

example. This command serves to restrict the amount of the data returned.

Example 3

The WDC database also permits registered users to store or replace short messages as part of their accessible information. This message facility is only available to the e-mail or telnet access methods. Messages are automatically output when *any* data item associated with the user's name is accessed. Messages cannot be searched for. Here Dr Maxwell deposits a message. Lucky fellow!

**user ukxxxfd0876 Maxwell**
**message**
*I am on sabbatical leave from April to December 1994 at the Department of Polynesian Studies,University of Tahiti. Please do not contact me unless you are desperate. Missing you already!*
**message_end**

Note that only Maxwell is able to insert or replace a message for his entry.

## *Gopher access mode*

*Gopher* is a search tool available on TCP/IP networks linked to the Internet. It is avaialable for computers ranging from Unix workstations to Apple microcomputers. Check the availability and usage of gopher with your network supervisor.

A gopher session on the WDC data base will use the address:
**diamond.iucr.ac.uk**

On entry, a menu will offer different database choices. Choose "World Directory 9th edition". A search sub-menu is provided. Fill in the data items to be searched for. The wild card character "*" may be used at the end of a string only.

The range of search options available with gopher will depend on your local interface connection. Searches may be performed on one, or many, fields. For example it may simply be on the name or using a combination of fields entered as follows:

**data_item1=value data_item2=value ...**

Appendix A lists the names of the searchable gopher data items.

## *WWW and Mosaic access modes*

*World Wide Web* is a general approach to accessing global databases which also provides facilities for gopher, telnet, or ftp access. A *WWW* user is automatically linked to the computer where the database exists.

*Mosaic* is a graphical interface which establishes links to data by clicking on areas of underlined text. The user must have a prior knowledge of how to link this

application to the primary server of the data. Check the availability and usage of WWW with your network supervisor.

To use Mosaic to connect to the WWW server in Chester, open the following URL (Uniform Rsource Locator):

**http://diamond.iucr.ac.uk/welcome.html.**

A menu will appear. Click on the line containing ????? This will open another sub-menu giving access to a variety of services in Chester. Choose "World Directory 9th edition" and click again. This will open a new window with a writable field for entering a query similar to that described for gopher above. The gopher service may be entered through Mosaic. For example,

**name=dupon* location=fr**

will return all names starting with "dupon" in France. The number of hits is limited to 25.

## *References*


Epelboin, Y. (1992a) *IUCr News Letters* 1, 2, 12-13.
Epelboin, Y. (1992b) *Acta Cryst.* **A48**, 949-954.
McMahon, B. (1993) *Acta Cryst.* **A49**, 222-225.
Hall, S.R. (1991) *J. Chem. Inf. Comput. Sci.* **31**, 326-333.
Spadaccini, N. & Hall, S.R. (1994) *J. Chem. Inf. Comput. Sci.* **34**, in press.


## *Appendix A*

## Searchable data items

**For e-mail and telnet access**

| | |
|---|---|
| _name_family | : family name |
| _name_other | : first names |
| _address_institution | : institution address |
| _address_postal | : preferred postal address |
| _phone | : phone number |
| _fax | : fax number |
| _telex | : telex number |
| _email | : electronic mail address |
| _degree_highest | : highest degree |
| _degree_date | : year where degree was conferred |
| _degree_field | : degree field |
| _birth_date | : year of birth |
| _current_position | : current position |
| _interests_key_words | : combination of keywords (Epelboin 1992b) |
| _interests_other | : unrestricted list of interests |
| _iucr_responsibilities | : IUCr or Regional Associate responsibilities |

**For Gopher and Mosaic access**

| | |
|---|---|
| family_name | : see _name_family |
| forenames | : see _name_other |
| year_of_birth | : see _birth_date |
| position | : see _current_position |
| iucr_responsibilities | : see _iucr_responsibilites |
| institution_address | : see _address_institution |
| postal_address | : see _address_postal |
| highest_degree | : see _degree_highest |
| degree_institution | : see _degree_institution |
| degree_date | : see _degree_date |
| degree_field | : see _degree_field |
| key_interests | : see _interests_key_words |
| other_interests | : see _interests_other |
| email | : electronic mail address |
| phone | : telephone number |
| fax | : facsimile number |
| telex | : telex number |

## *Appendix B*

### Location codes

ISO 3166 standard country codes

| | |
|---|---|
| am | Armenia |
| ar | Argentina |
| at | Austria |
| au | Australia |
| az | Azerbaijan |
| be | Belgium |
| bg | Bulgaria |
| br | Brazil |
| ca | Canada |
| ch | Switzerland |
| cl | Chile |
| cn | China |
| cu | Cuba |
| cz | Czech Republic |
| de | Germany |
| dk | Denmark |
| ee | Estonia |
| eg | Egypt |
| es | Spain |
| fi | Finland |
| fr | France |
| gb | United Kingdom |
| gr | Greece |
| hk | Hong Kong |
| hr | Croatia |
| hu | Hungary |
| id | Indonesia |
| ie | Ireland |
| il | Israel |
| in | India |
| it | Italy |
| jp | Japan |
| kr | Korea |
| lv | Latvia |
| md | Moldova |
| mm | Myanmar |
| mx | Mexico |
| my | Malaysia |
| ne | Niger |
| nl | Netherlands |
| no | Norway |
| nz | New Zealand |
| pk | Pakistan |
| pl | Poland |
| pt | Portugal |
| ro | Romania |
| ru | Russia |
| se | Sweden |
| sg | Singapore |
| si | Slovenia |
| sk | Slovakia |

| | | | |
|---|---|---|---|
| th | Thailand | uz | Uzbekistan |
| tm | Turkmenistan | ve | Venezuela |
| tr | Turkey | vn | Vietnam |
| tw | Taiwan | yu | Serbia |
| ua | Ukraine | za | South_africa |
| us | United States | | |
| uy | Uruguay | | |

## *Appendix C*

### Query language Summary

**user** ***<security code> <family name>*** [mandatory]
This sequence is the first entry by a user. The registered security code and user family name are entered as a unique combination.

**access** ***<access level>*** [*general, journal, national, regional, super*]
The access code specifies the current search and hit permissions. The codes a user may enter depend on the access class assigned to the user in the WDC registration letter. The default code is "*general*".

**limit** ***<upper hit limit>*** [optional]
This sets the maximum number of successful matches (i.e. "hits") for the current search. This number may not exceed the assigned limits associated with the access class, or the accumulated number of hits for the current calander month.

**location** ***<location code>*** [optional]
This command restricts searches according to geographical location. The country or regional codes are listed in Appendix B. The default is "world".

**message** [optional]
This command starts a text message to be inserted into the user information. If any information for this user is accessed this message will be listed in the search output. The message must be closed by **message_end**.

**if** ***<conditional search string>*** [optional]
This command is used to restrict the information output according to specific matching of text entries. The construction of the *conditional search string* is

***<data item name> <operator> <value>***

The permitted search *operators* are:

| | |
|---|---|
| **==** | "exact match" |
| **~=** | "containing string" |
| **!=** | "NOT exact match" |
| **!~** | "NOT containing string" |

Sequences of the conditional search string may be concatenated with the Boolean operators:

| | |
|---|---|
| **&&** | "and" |
| \|\| | "or" |

The wild card character "*" may be used in data names only; not in the search values. The default is "all entries".

**want** ***<data items to be returned>*** [optional]
This specifies which data items are to be returned. The default is all items.

**ignore** ***<data items NOT to be returned>*** [optional]
This specifies those data items which should not be returned. The default is null. The "want" and "ignore" commands should not be used in the same search.

**search** [optional]
This command starts a search which is based on the preceding query parameters. Multiple searches may be performed in a single session. Unless search conditions are re-specified, those used in preceding searches will apply.