

# MSD Search Database Practical Session

EMBL-EBI, Wellcome Trust Genome Campus

This is a demonstration of a stand-alone replica version of the MSD Search Database (MSDSD) in mySQL, together with tools and examples about how to use it. All the software together with the demo database is self-contained and wrapped up in a single CR-ROM disc that may be installed and used directly in any Windows or Linux machine. There are no other special pre-requisites apart from 4.5 GB of free disk space.

This is a cut-down version of the MSDSD in the amount of data that it includes and this was imposed by the need to have a standalone mySQL database of a size that is practical for demonstration purposes (when compressed fits in a CD-ROM). So while the full width of MSDSD is included (100 tables - 1948 columns) instead of the 22846 PDB entries that MSDSD currently includes, this has data for

- 3096 entries that are included in the SCOP list of representative entries for most tables except
- 244 entries in the SCOP list with resolution < 1.5 on the residue level (tables RESIDUE, RESIDUE\_DATA etc)
- 83 entries in the SCOP list with resolution < 1.2 on the atom level (tables ATOM, ATOM\_DATA etc)

You may get a copy (subject to licensing restrictions) of this MSD-demo CD-ROM disc and follow these steps in order to repeat the MSD tutorial later. The purpose of this CD-ROM is for demonstration and educational purposes only. Please read carefully the licence.txt file before start using the CD-ROM disc.

This CR-ROM includes

- a) Java 1.4 preinstalled for windows and linux
- b) mySQL 4.0 preinstalled for windows and linux
- c) The QueryForm Database Tool
- d) Rhino 1.5 from Mozilla
- e) mySQL ODBC drivers 02.50.38, setup for windows
- f) Jakarta Tomcat 4.1 servlet container
- g) The MSDSD (MSD search database) a cut-down demo version preloaded in MySQL
- h) The MSD-mine servlet web application setup to run locally and access the local mySQL database.
- i) Example files and scripts
- j) The standalone static MSDSD data warehouse documentation

## 1) Install the MSD-demo CD-ROM disc

This may have been done already for you in advance, in order to save time so ask your tutor before you start.

Follow the steps in the install.txt file of the CD-ROM disc and run the install.bat script. This will unzip the msddemo.zip file on drive C:\ and create the C:\msddemo directory.

Test that everything is OK. Open "My Computer" or the Windows "Explorer" and check that C:\msddemo exists and is populated with files and directories

## 2) Configure MSD-demo

Run the script C:\msddemo\bin\config.bat. This will create a virtual drive O: and will copy the mySQL ini files in the appropriate locations. You will have to rerun this script after each reboot of the host in order to re-create the virtual drive O:

Test that everything is OK. Open "My Computer" or the Windows "Explorer" and check that the virtual drive O: exists. Also check that the file C:\my.cnf (mySQL configuration file) exists

## 3) Start the mySQL database

Run the script O:\bin\startdb.bat. Leave the mySQL server window open

040114 10:37:02 InnoDB: Started  
mysqld: ready for connections

#### 4) Use the mySQL command prompt

This is the default mySQL prompt for giving queries and retrieving results. Run O:\bin\mysql.bat  
Welcome to the MySQL monitor. Commands end with ; or \g.  
Your MySQL connection id is 10 to server version: 4.0.1-alpha-max-debug  
Type 'help;' or '\h' for help. Type '\c' to clear the buffer.  
mysql>

Now you may try some SQL statements like  
Get the number of assemblies in the MSDSD

```
mysql> SELECT COUNT(1) FROM ASSEMBLY;
```

```
+-----+  
| COUNT(1) |  
+-----+  
|      4529 |  
+-----+
```

1 row in set (0.54 sec)

Get information for the biggest chains and the accessible surface area of their assemblies

```
mysql> SELECT ACCESSION_CODE AS PDB, ASSEMBLY_SERIAL AS  
ASSEMBLY, CHAIN_CODE,  
          ASSEMBLY_TYPE, NUM_RESIDUES AS RESIDUES, ASA  
FROM CHAIN WHERE NUM_RESIDUES > 1200 AND ASA IS NOT NULL  
ORDER BY NUM_RESIDUES DESC;
```

```
+-----+-----+-----+-----+-----+-----+  
| PDB   | ASSEMBLY | CHAIN_CODE | ASSEMBLY_TYPE | RESIDUES | ASA      |  
+-----+-----+-----+-----+-----+-----+  
| 1jj2  |          | 0          | 30MERIC       | 2922    | 364565.38 |  
| 1i50  |          | A          | DECAMERIC     | 1733    | 74247.75  |  
| 1fo4  |          | B          | DIMERIC       | 1332    | 47879.93  |  
| lepw  |          | A          | MONOMERIC     | 1290    | 0          |  
| lkek  |          | B          | DIMERIC       | 1231    | 51390.2   |  
| 1i50  |          | B          | DECAMERIC     | 1224    | 54857.38  |  
+-----+-----+-----+-----+-----+-----+
```

12 rows in set (1.07 sec)

Find the entries with the most secondary structure "turn" elements that start with an Alanine

```
mysql> SELECT ACCESSION_CODE, COUNT(1) AS NUM_TURNS  
FROM TURN  
WHERE NON_ASSEMBLY_VALID='Y' AND RES_1_CHEM_COMP_CODE='ALA_LL'  
GROUP BY ACCESSION_CODE  
HAVING NUM_TURNS > 8;
```

```
+-----+-----+  
| ACCESSION_CODE | NUM_TURNS |  
+-----+-----+  
| lezg           | 10        |  
| lgoi           | 11        |  
| lhbn           | 13        |  
| ljcl           | 12        |  
+-----+-----+
```

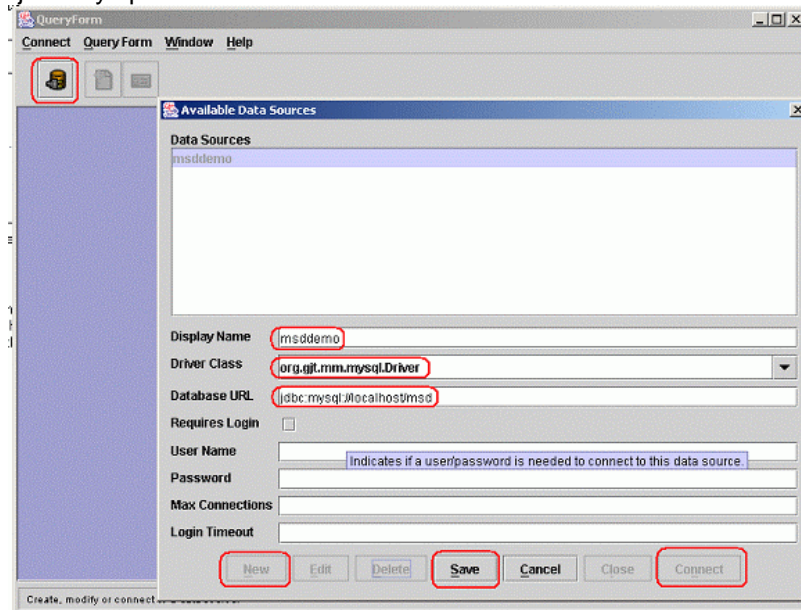
4 rows in set (0.17 sec)

You may write SQL queries that use any other of the 100 tables and 1948 tables, and even do table joins or other "group by" expressions etc.

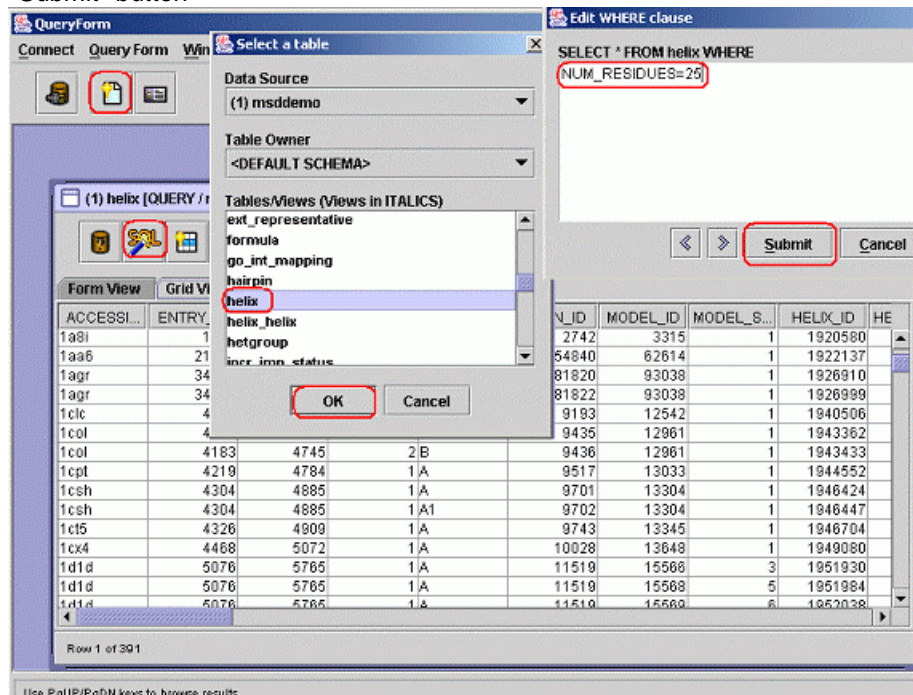
#### 4) Use the QueryForm tool

This is a freeware GUI application that allows the execution and result view of SQL statements to any database through JDBC. Run O:\bin\qform.bat  
Click on "Connect:Data Source ..." icon and press the "New" button.

Select "org.gjt.mm.mysql.Driver" as the "Driver Class", "msddemo" as the "Display Name" and "jdbc:mysql://localhost/msd" as the Database URL. Press "Save" and then "Connect"



Click on the Menu "Query Form:New Query Form", choose the "helix" table and press "OK" Select the "Grid View" click on the "SQL" icon, give "NUM\_RESIDUES=25" and press the "Submit" button



You may also use any other table in the list an perform more queries



