

Queensland Museum

- Source-data description
- Exporter components description
- People
- Prerequisites and notes
- Recommended modifications procedure
- Running an export

Source-data description

Data are managed in the primary QM collection management system, a Vernon system and database.

A snapshot of a chosen subset of records is hosted externally to QM infrastructure, within a MySQL database. The records that are included in the snapshot are controlled by QM's source-system.

Information from each record in the snapshot is predominantly stored in 100 fields, named `text1..text100`; the source content stored within these fields is also defined via configuration in QM's source-system. The snapshot also stores multimedia in some instances, with one or more images being cross-referenced to one or more specimen records.

The frequency of record updates is unknown, but is believed to be near real-time, with the snapshot record being updated automatically as the source data change (a potential therefore exists for a delta-export to piggyback off this process).

Should the content stored in the snapshot fields change (*text1..text100*), an update to the logical mapping (hosted in mysql view `vw_qmdwc` defined in [qmdwc_view.sql](#)) between these and the Darwincore terms would be required.

Exporter components description

There are two main parts to the exporter: the export components and the exporter update.

The first export component is a bash shell script, [/qmexport/qmdwc.sh](#), which is the entry point for running an export – this script connects to the mysql database, combines the results of queries to extract data into a darwincore-csv format, and uploads the result to the [upload.ala.org.au](#) sftp server.

The second export component is the mysql query, [/qmexport/qmdwc_export.sql](#), which is called by [qmdwc.sh](#) and extracts data from the view. This query will reliably extract data in whatever the contents of the view, honouring any changes without needing an update. Output is in two files: a header row [/tmp/qmdwc_header.csv](#) and data rows [/tmp/qmdwc_data.csv](#)

Note: a process must then parse and combine these two files for data to be made available in a single file (see [qmdwc_export.sql](#) for more details).

The third export component is the mysql view `vw_qmdwc`, which hosts the logical mapping between the snapshot schema and darwincore. This main view is supported by `vw_qmavlist`, which builds the list of image urls. Both of these views are defined in [/qmexport/qmdwc_view.sql](#)

If the content configuration in the snapshot fields change, then the logical mapping between these (*text1 .. text100*) fields (`vw_qmdwc` and [qmdwc_view.sql](#)) must also change. If this happens, a full export will be required as all data will be affected by the change in configuration.

Note: the content configuration can only be defined by QM, from within QM infrastructure.

Finally, [/qmexport/qmupdate.sh](#) exists to pull down latest copies of all these files and trigger [qmdwc_view.sql](#) to update the views. This script will first connect to the sftp upload server, to download the files, then to the local mysql server to update the views. It will finish by issuing the mysql command `describe vw_qmdwc`; which will show the dwc terms in the logical dwc mapping.

People

Paul Avern – paul.avern@qm.qld.gov.au – +61 7 3840 7711

Paul Rowe – paul@vernonsystems.com – +(649) 815-5599

Prerequisites and notes

Access to the Vernon-hosted snapshot is required, described previously in *source-data*, and must be teed up with QM and Vernon systems. In addition to this, minimum software needed is 1) a SSH terminal, such as Putty; and 2) a SFTP client, such as WinSCP

As this server is managed by third parties, you should keep a log of your session: (putty → rightclick title bar → copy all to clipboard) and paste into a separate text file – **note**: this file should be kept securely, as there are passwords echoed in parts of the session's output.

You should not change the view without updating the corresponding view definition file (*qmdwc_view.sql*) on the sftp server, as well as in the */qmexport/* directory on the snapshot server. This will ensure an earlier version is not inadvertently 'rolled back'.

Finally, if the logical mapping changes then the completeness model should also be updated to reflect this:
<http://goo.gl/Lwmu5>

Note: *text in italics* indicates command-line input or output while CAPITAL_BLUE words act as place-holders for user accounts and passwords.

Running an export

- 1 – log in to the vernon snapshot server: *VERSNAP_USER@VERSNAP_IP_ADDY* (*pwd: VERSNAP_PWD*)
- 2 – move to the /qmexport directory (*cd /qmexport*)
- 3 – trigger an export (*./qmdwc.sh*) (1st mysql *usr: MDB_USER* *pwd: MDB_PWD*) (2nd sftp *pwd: SFTP_QMPWD*)
- 4 – look for *#./qmdwc.sh#23:20:57# finished* or similar in the output, indicating the script has finished
- 5 – check a file with a recent timestamp has been moved to history, **note**: server is UTC (*ls -l sftp_history/*)
- 6 – trigger a listing of the sftp directory (*echo "ls -l" | sftp SFTP_QM@upload.ala.org.au*) (*pwd: SFTP_QMPWD*) and look for the same file identified in the previous step; this will indicate that the file was uploaded successfully

The exporter bash shell script (*qmdwc.sh*) will always bundle the latest copies of all exporter files, as well as use *mysqldump --no-data* to generate a sql dump of the schema, which may aid in uncovering any discrepancies that arise.

Recommended modifications procedure

- 1 – make changes to files in the *SFTP_QM@upload.ala.org.au/_exporter/* directory (*pwd: SFTP_QMPWD*)
- 2 – log in to the vernon snapshot server: *VERSNAP_USER@VERSNAP_IP_ADDY* (*pwd: VERSNAP_PWD*)
- 3 – move to the /qmexport directory (*cd /qmexport*)
- 4 – trigger an update (*./qmupdate.sh --yes*)
(1st sftp *pwd: SFTP_QMPWD*) (2nd mysql *usr: MDB_USER* *pwd: MDB_PWD*)
- 5 – check your new content has been downloaded from sftp (*cat <yourfile.ext>*)