Australian Museum

Source-data description

Data are managed in the primary AM collection management system, a Ke EMu system and texpress database. AM's public records are stored in a duplicate instance, with institution-private records excluded. This snapshot is updated periodically by AM staff.
Note: the exporter assumes private records aren't available in the snapshot, so doesn't filter for this purpose.

Exporter components description

There are five parts to the exporter: the main script, the disciplines-list, the sub-script, the data mapping scripts and the update script.

The first export component is a bash shell script, **/amexport/dwcdm2.sh**, which is the entry point for running an export – this script prepares the export directory, reads in **disciplines-list** and calls the sub-script **dwcdm2dsx.sh** for each non-comment line (no leading #), bundles the export on completion and sends to specified servers using sftp.

The second export component is the text file, **/amexport/disciplines-list**, which controls the behaviour of the main script **dwdm2.sh**. Disciplines matching the **CatDiscipline** field should be entered here, one per line. Comment lines (beginning with #) and blank lines are ignored. Comment out disciplines to do a partial export. If you delete or rename this file, **dwcdm2.sh** will rebuild it from the database.
Note: this is a costly operation (roughly 2 hours) and no subsequent exports will occur, to allow for any unwanted disciplines to be excluded by deletion or comment.

The third export component is a bash shell script, **/amexport/dwcdm2dsx.sh**, which is called by **dwcdm2.sh** for each line in the disciplines-list, and handles exporting the full list of current id's, as well as the partial or full export depending. Scripts **ozdc_full.awk** and **ozdc_id.awk** are called by **dwcdm2dsx.sh**, and handle mapping between an emu export and a darwincore csv. The output file **DISCIPLINE-dwcid.csv** is converted by **ozdc_id.awk** while **DISCIPLINE-dwcdata.csv** is converted by **ozdc_full.awk**.

The fourth export component are the awk scripts, **/amexport/ozdc_full.awk** and **/amexport/ozdc_id.awk**. These scripts handle the mapping between an emu export and a darwincore csv – they are called on by **dwcdm2dsx.sh** to convert data inline before output csv files are written by **dwcdm2.sh**.

Finally, **/amexport/amupdate.sh** exists to pull down latest copies of all these files. This script will connect to the sftp upload server then download the files in the sftp **/_exporter** directory to the current working directory when this script is run (see: *pwd*).

People (in order of preference)

Mike Elliott - Michael.Elliott@austmus.gov.au – +61 2 9320 6345
Paul Flemons – Paul.Flemons@austmus.gov.au – +61 2 9320 6343

Prerequisites and notes

The following procedures require terminal access from within AM's network (eg: putty).

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

The AM dmz-infrastructure doesn't allow dns lookups, so the IP address for the sftp server is hard-coded in the **dwcdm2.sh** script, as well as the following procedures. You should check from a machine with dns lookup permissions (e.g. a machine on the internal network) to ensure that **upload.ala.org.au** still resolves to **175.41.168.229** (command: *nslookup upload.ala.org.au*).

Servers are 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.

Finally, if the logical mapping changes then the completeness model should also be updated to reflect this: http://goo.gl/0NLPR

Running an export

1 – start a new terminal session within the AM internal network

2 – log in to the primary emu server (*usr:* EMU_ADMIN *pwd:* EMU_PWD)

3 – ssh to the web-facing snapshot (*ssh* DMZ_EMU_IP_ADDY)

4 – move to the **/amexport** directory *(`cd /amexport`)*

6 – trigger an export *(./dwcdm2.sh)*

7 – look for *`#./dwcdm2.sh#23:20:57# finished`* or similar in the output, indicating the script has finished

8 – check a file with a recent timestamp has been moved to history *(ls -l sftp_history/)*

9 – trigger a listing of the sftp directory *(echo "ls -l" | sftp* SFTP_AM@175.41.168.229*)* (*pwd:* SFTP_AMPWD) 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 **dwcdm2.sh** will always bundle script output, the latest copies of all exporter files, as well as use texpress commands to generate a dump of the schema (see the script for more details), which may aid in uncovering any discrepencies that arise.

Recommended modifications procedure

1 – make changes to files in the SFTP_AM@175.41.168.229/_exporter/ directory (*pwd:* SFTP_AMPWD)

2 – log in to the primary emu server (*usr:* EMU_ADMIN *pwd:* EMU_PWD)

3 – ssh to the web-facing snapshot (*ssh* DMZ_EMU_IP_ADDY)

3 – move to the /amexport directory *(cd /amexport)*

4 – trigger an update *(./amupdate.sh --yes)* (*pwd:* SFTP_AMPWD)

5 – check your new content has been downloaded from sftp *(cat* <yourfile.ext>*)*

Recommended installation procedure

1 – log in to the primary emu server using the emu admin account (*usr:* EMU_ADMIN *pwd:* EMU_PWD)

2 – ssh to the web-facing instance (*ssh* DMZ_EMU_IP_ADDY)

3 – test texpress commands are available by running the following commands (with expected output):
texlist -l (listing of all texpress tables, similar to: TABLE_NAME *(user) 7654321 records 98.7%*)
echo describe TABLE_NAME | *texql -R* (description of the chosen table from the *texlist* output)

4 – create the /amexport directory *(sudo mkdir /amexport)*

5 – change owner of the /amexport directory to the emu admin *(sudo chown -Rc* EMU_ADMIN */amexport)*

6 – download the files *(echo "get _exporter/*" | sftp* SFTP_AM*@175.41.168.229)* (*pwd:* SFTP_AMPWD)

7 – ensure the main script is executable *(chmod +x dwcdm2.sh)*

8 – ensure the sub-script is executable *(chmod +x dwcdm2dsx.sh)*

9 – ensure the update script is executable *(chmod +x update.sh)*

10 – check all component files are present, with appropriate permissions *(ls -l /amexport/)*