South Australian Museum

Source-data description
Data are managed in the primary SAM collection management system, a Ke EMu system and texpress database.

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, **/samexport/dwc_spc.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 **dwc_spsub.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, **/samexport/disciplines-list**, which controls the behaviour of the main script **dwc_spc.sh**. Disciplines matching the **CatCollectionName** 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, **dwc_spc.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, **/samexport/dwc_spsub.sh**, which is called by **dwc_spc.sh**, 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 **dwc_spsub.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, **/samexport/ozdc_full.awk** and **/samexport/ozdc_id.awk**. These scripts handle the mapping between an emu export and a darwincore csv – they are called on by **dwc_spsub.sh** to convert data inline before output csv files are written by **dwc_spc.sh**.

Finally, **/samexport/samupd.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 **/_dmscripts** directory to the current working directory when this script is run (command: *pwd*) - see modification and installation procedure defined later.

People
Robert Morris - Robert.Morris@samuseum.sa.gov.au – +61 (0)8 82077455

Prerequisites and notes

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

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.

If modifications are required, a full export should be run afterwards; data should replace existing Atlas records. **Note:** an update can only be run by SAM, from within SAM infrastructure.

Finally, if the logical mapping changes then the completeness model should also be updated to reflect this: http://goo.gl/uVI9x

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 – start a new terminal session within the SAM internal network

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

3 – move to the **/samexport** directory *(`cd /samexport`)*

4 – trigger an export *(./dwc_spc.sh)*

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

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

7 – trigger a listing of the sftp directory *(echo "ls -l" | sftp* SFTP_SAM@upload.ala.org.au*)* (*pwd:* SFTP_SAMPWD) 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 *(dwc_spc.sh)* will always bundle the latest copies of all exporter files, as well as using texpress commands to generate a sql dump of the schema (see script for more details), which may aid in uncovering any discrepencies that arise.

Recommended modifications procedure

1 – make changes to files in the SFTP_SAM@upload.ala.org.au/_dmscripts/ directory (*pwd:* SFTP_SAMPWD)

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

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

4 – trigger an update *(./samupd.sh --yes)* (*pwd:* SFTP_SAMPWD)

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

Recommended installation procedure

1 – connect to the primary emu server using putty or similar

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

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 /samexport directory *(sudo mkdir /samexport)*

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

6 – download the files *(echo "get _dmscripts/*" | sftp* SFTP_SAM*@upload.ala.org.au)* (*pwd:* SFTP_SAMPWD)

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

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

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

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