Australian National Fish Collection

Source-data description

Data managed in the primary ANFC collection management system, a texpress database.

The records that are included in the export are controlled by a report that is run (currently ad-hoc) by the collection manager. Data included in this report are then parsed and converted to Darwincore csv format.

Exporter components description

The main export component is a pentaho d.i. (aka kettle) transformation, **anfc-csv.ktr**, which is a prototype data migration utility. This transformation script takes the results of the texpress report, provided by the collection manager. These data are in a text file, currently delimetered with the sequence: +|

The transformation script converts data from the source system into darwincore csv form. A more complete description for using this file is available at:
http://code.google.com/p/ala-datamob/source/browse/trunk/biodomains/fcig-ozcam/anfc/anfc-dwc+pentaho+transfo

Recommended modifications procedure

None – this transformation is a prototype for data analysis and discovery; it relies on a support-package with a steep learning curve (pentaho). There are no plans at this stage to develop it further, rather, to port the proven logic to a perl + bash solution for the linux texpress backend.

Before the effort to port the logic happens, this script would prove useful if an updated export is required, or additional data must be added quickly.

A completeness model also exists to guide further development: *http://goo.gl/tKRzb*

People (in order of preference)

Al Graham – alastair.graham@csiro.au – +61 3 6232 5321

Prerequisites

Pentaho data integration (aka kettle) –
http://sourceforge.net/projects/pentaho/files/Data%20Integration/4.2.1-stable/