TRiCYCLE Manual

Peter Brewer
Daniel Murphy

TRiCYCLE Manual

by Peter Brewer and Daniel Murphy

Acknowledgements

Chapter 1. What is TRiCYCLE

TRiCYCLE is a universal dendrochronology file format converter. It currently has support for reading and writing 17 different file formats:

- Belfast Apple

- Belfast Archive

- Besancon

- CATRAS

- Comman Separated Values (CSV)

- Corina Legacy

- Excel

- Heidelberg

- Nottingham

- Sheffield

- TRIMS

- TRiDaS

- Topham

- Tucson

- Tucson Compact

- VFormat

- WinDENDRO

TRiCYCLE extracts both data and any metadata present in files and converts them to the Tree-Ring Data Standard (TRiDaS) data model. As TRiDaS is capable of representing the full range of dendro data and metadata, it is then possible to write out the file to any one of the supported formats.

Key features of TRiCYCLE are:

- Seamless support for units where possible

- Interpretation of all metadata

- Handling of different charactersets and line feeds from different operating systems

- Comprehensive warning and exception system which provides detailed feedback when errors are detected in files

For a complete discussion of TRiCYCLE and its underlying libraries please see [Brewer (in review)]

Chapter 2. Installation

TRiCYCLE is a Java application and so can be installed on any modern operating system. To make installation more familiar though we have packaged it up into native installers for Windows, Mac OSX and Linux. Download the relevant package for your operating system from the SourceForge [http://tridas.sf.net] website.

Windows Run the setup program and follow the steps. The program will be installed to your hard disk and shortcuts added to your start menu. If you do not have Java installed on your system or you do not have the required version the installer will provide you with assistance to do so.

MaxOSX Open the .dmg file in Finder. Drag the TRiCYCLE.app file into your Applications folder, or wherever you'd like it installed. MacOSX comes pre-installed with Java so there is no need to install it separately.

Linux An Ubuntu deb package is provided for TRiCYCLE. Install this package with your standard package manager (e.g. dpkg). It includes information on all the dependencies required including Java therefore it should install everything you need if you are running Ubuntu. On other distributions you may find that the dependencies do not install automatically and the TRiCYCLE installation fails. If this is the case you will have to install the following dependencies manually:...

Chapter 3. How to use TRiCYCLE

The first step is to specify the format of your input file(s). Choose your input file format from the pull down list on the File List page. Then select the file(s) that you'd like to convert by clicking on the browse button. You can add files one at a time or in bulk. Once you have your list of files prepared you can then go to the Convert tab.

On the Convert page, the next thing you need to do is select the desired output format from the drop down list. Next, press the convert button and after a short delay your files should appear in the list below. At the bottom of the window, you'll see a summary showing how many files were processed, how many of the convertions failed and how many were converted successfully but with warnings.

The results of the conversion are shown in a tree view on the convert tab. Files that failed to convert are highlighted by a red cross along with a message explaining what went wrong. Each file that convered successfully and with no warnings is shown with a green tick icon below which are shown the output file or files. Depending on the input and output formats chosen, a input file may be converted into one or more output files as some formats can store just one series while others can store multiple series.

If there were warnings produced during the conversion process, then the file will be marked with a orange exclamation icon. Warnings can be associated with either the process of reading the input file, or writing the output file. They can also be related to a single series within the file or the input file as a whole. The warning messages are displayed to illustrate the context of the warning. Longer warnings will scroll of the edge of the window, but if you hover your mouse over the warning a tooltip will show the entire message.

If you would like to preview the result of the conversion process you can double click on the output files and they will be displayed in a text viewer. This option is not, however, available for binary formats such as CATRAS and Excel.

Once you are happy with the results of the conversion you can save the files permanently to disk by pressing the save button. It will offer you the option of specifying which folder to save the files to.

Chapter 4. Options

The options panel is available from the file menu. It is split into two sections: Reader config and Writer config.

Character sets

The character set can be set for both the file being read and the file being written. The character set is the system for pairing computer character codes with the character glyphes that we read. The widely used standard was originally ASCII, but this does not include diacritic characters, and characters specific to certain languages. There have since been many character encodings proposed (e.g ISO 8859-1 for Western Europe and ISO 8859-7 for Greece) as well as some that are specific to Windows and Mac operating systems (e.g. Windows-1252 and MacRoman). The character set this is becoming most widely used is the Unicode UTF-8. This is capable of representing all 107,000+ distinct characters while remaining backwards compatible with ASCII for the 128 characters that it ASCII is able to represent.

If an incorrect character encoding is used to interpret a file, normally the majority of characters will display correctly (where the character sets share the same encodings) but more unusual characters will be displayed incorrectly - typically square boxes or question marks.

TRiCYCLE can using the NIO package to attempt to automatically detect which encoding a file is in. Unfortunately, there is no full-proof way to do this so by default, this feature is turned off. If you are having problems with character encodings you may like to choose 'Automatic' in the charset box if you have no idea what character encoding your file is in.

The character encoding is set to the default for the operating system you are running. For instance on MacOSX this will be MacRoman and for Windows it will be Windows-1250. If you know your input file is in a different encoding you should set it in the input charset box. If your output file needs to be read on an operating system other than the one you are currently running, then you may like to override the writer charset. Please note that for certain writers the character set used is part of the file specification (e.g. TRiDaS must be UTF-8). In this case your choice will be ignored.

The final complication with regards character sets is the line feed character(s). For historical reasons different operating systems use different characters to represent a new line. Depending on the software that is used to read a file, this can cause problems. TRiCYCLE itself will automatically adapt to files with any type of line feed characters so reading files in TRiCYCLE will never be a problem. When writing out files, TRiCYCLE will use the default line feed for the operating system you are running on unless you choose a platform specific character set. For instance if you run TRiCYCLE on Windows and choose a MacRoman writing charset, TRiCYCLE will use Mac style line feeds.

Metadata editor

TRiCYCLE works by reading in a data file and translating it into the TRiDaS data model. TRiDaS has a rich array of fields to represent all manor of dendro data and metadata. Although most of these are optional, the TRiDaS specification requires that a handful of these are always filled in. Unfortunately many of the legacy data formats do not contain information for these mandatory fields, therefore TRiCYCLE must fill these with default values. The metadata editor enables you to override these default values.

Clicking on the reader metadata editor button in the options window will give a table of all the metadata fields that will be set automatically by TRiCYCLE along with their current values. You can change most of these with the exception of those that are required to be a controlled vocabulary. These will require a more complicated interface which we haven't had time to implement yet. The third column in the editor is

a tick box to specify whether the value is overriding or not. If ticked, the value specified in this editor will be used regardless of whether a value can be extracted by TRiCYCLE from the input datafiles.

An identical editor is available for the writer. These are the default values used by the writer code for your chosen output format. For instance, TRiDaS does not require that a start year field be set (as in the case of relatively dated series), whereas some output formats do require such a field. If an input file does not contain start year information then some writers need to know which default value for start year to use. Like for the input metadata editor, you can set fields to 'overriding' which means they will be used regardless of whether this information is available in the input dataset.

Naming convention

Some file formats can contain just one data series while other can contain many. When converting from a multi-series format to a single series format this means that one input file is converted to multiple output files. The naming convention is used to determine how to name the output files. The naming convention relates to the filename itself and not the file extension. The file extension is specific to the output format chosen (e.g. Heidelberg files are .fh and TRiDaS files are .xml).

Numerical   This is the default naming convention. It uses the name of the input data file and appends an incrementing number if more than one output file is produced.

UUID   This gives all output files a random named based on Universally Unique Identifiers (UUIDs). This is a 36 character hexidecimal code which due to the astronomically large number of possible combinations is garanteed to be universally unique. A typical filename will look like: 550e8400-e29b-41d4-a716-446655440000.

Hierarchical   This uses the hierarchical structure of the TRiDaS data model to provide a meaningful name for the output file. It joins together the title of each entity in the file beginning with the project name through to the series name. For files that contain multiple series, the name will contain details of all the entities shared by all the series in the file. For example, if a file contains several series from the same sample, then the file name will be projectTitle-objectTitle-elementTitle-sampleTitle. If the file contains several series from different samples of the same object, then the file would be projectTitle-objectTitle. If multiple output files end up with the same name then like the numerical convention described above, the files will have an incremental number appended to the end. Unfortunately, most input data files do not contain rich name information so files end up being called unnamedProject-unnamedObject-unnamedElement etc. This convention is therefore more appropriate when converting from TRiDaS to other formats.

Chapter 5. Help and more information

The best place to start is through the TRiDaS website (www.tridas.org [http://www.tridas.org]) and the Dendro Data Standards forum. The forum is a email list for the discussion of TRiDaS and other dendro data standards issues. It is open to all to join by emailing Peter Brewer [mailto:p.brewer@cornell.edu].

TRiCYCLE is an open source product therefore we are very pleased to welcome anyone that would like to assist in its development. This obviously includes programmers, but also people willing to help with documentation and translations too. To find out more information please contact Peter Brewer [mailto:p.brewer@cornell.edu].

Bibliography

R Aniol. 1983. *Tree-ring analysis using CATRAS*. Dendrochronologia. 1. 45-53.

P Brewer, K Sturgeon, L Madar, and S Manning. 2010. *A new approach to dendrochronological data management*. Dendrochronologia. 28. 131-134.

[Brewer (in review)] P Brewer, D Murphy, and E Jansma. in review. *TRiCYCLE: a universal dendro data conversion tool*. Tree-Ring Research.

A Bunn. 2008. *A dendrochronology program library in R (dplR)*. Dendrochronologia. 26. 115-124.

R Holmes. 1983. *Computer-assisted quality control in tree-ring dating and measurement*. Tree-Ring Bulletin. 43. 69-78.

R Holmes. 2001. *Dendrochronology Program Library (DPL)*. The University of Arizona, Tucson..

E Jansma, P Brewer, and I Zandhuis. 2010. *TRiDaS 1.1: The tree-ring data standard*. Dendrochronologia. 28. 99-130.

B Knibbe. 2008. *PAST4 - Personal Analysis System for Tree Ring Research. Software version 4.5*. SCIEM.

F Rinn. 2008. *TSAP-Win software*. Rinntech, Heidelberg.

J Speer. 2010. *Fundamentals of Tree Ring Research*. University of Arizona Press.

I Tyers. 1999. *Dendro For Windows program guide*. ARCUS Report 500.