Appendix: DateValue Input Type

2011-05-09

Overview

The DateValue time series file format can be used to store one or more time series of consistent time interval. The format has been developed to facilitate data management and processing with the TSTool software. The example below shows the format of the file. Important comments about the file format are:

- The file is divided into a header section with time series properties (top) and data section (bottom). Comments can occur anywhere in the file and are lines that start with #.
- The default delimiter between property columns and data columns is a space. Use the `Delimiter` property to reset the delimiter (e.g., to a tab). Adjacent delimiters WILL NOT be merged into one column. **It is recommended that a single space be used as the delimiter between header values and data columns.**
- If not specified, many of the header properties will be set to reasonable defaults as data are read by software such as TSTool. However, as much information as possible should be specified to allow complete time series handling. Header information is displayed by applications like TSTool to allow selection of time series before the data section is read.
- Properties names are case-independent.
- The `TSID`, `Start`, `End`, and `Units` properties are important for basic time series handling.
  - The interval part of the TSID is used to determine how memory should be allocated for data.
  - The `Start` and `End` values are used to allocate memory for regular interval time series. Dates associated with data values are used to allocate memory for irregular interval time series.
  - For regular interval time series, if data lines between the start and end dates are omitted, the unspecified values are set to the missing data value for the time series.

```
# DateValueTS 1.1 file
#
# This is a sample of a typical DateValue minute time series.  An
# example file is as follows and conforms to the following guidelines:
#
# * Comments are lines that start with #.
# * Applications often add a comment section at the top indicating how the
#   file was created
# * Any line that starts with a number is assumed to be a data line.
# * Time hours should be in the range 0 to 23 (an hour of 24 will be
#   converted to hour 0 of the next day).
# * If a time is necessary, the date/time may be separated by a space, T, :, or
#   @.  If a space is used, use date and time column headings.
# * If multiple time series are written, header variables are delimited with
#   space or tab characters.  Data are delimited as per the Delimiter property.
# * Time series are required to have the same data interval.
# * Header property values and column headers can be enclosed in double quotes
#   if the data contain spaces.
# * Missing data can either be coded as the missing data value or no value.
# * Missing records will result in missing data being used when read.
#
# The following header variables are recognized.  This information can be
# used by software.
Version = 1.1                   # Optional.  File format version
                                # (to handle format changes)
Delimiter = " "                 # Optional.  Delimiter for property and data lines
```

```
                                    # (default is space)
NumTS = 2                           # Optional.  Number of time series in file
                                    # (default is 1)
TSID = "XXX.USGS.Streamflow.15MINUTE" "YYY.USGS.Streamflow.15Minute"
                                    # Required.
                                    # List of time series identifiers in file
                                    # Location.Source.DataType.Interval.Scenario
                                    # Do not include input type and name in identifier
Alias = "XXXX-Streamflow" "YYYY-Streamflow"
                                    # Optional – assign the time series alias
SequenceNum = 1950 1951             # Optional – used with time series ensembles.
                                    # Indicates the year for each trace
Description = "Flow at XXX" "Flow at Y"
                                    # Optional.  Description for each time series.
DataFlags = true false             # Optional.  Indicates whether data flags for the
                                    # data values are provided.  If flags are used,
                                    # the data value column a time series with data
                                    # flags is followed by a column for the data flag.
                                    # Surround the flags by "" if a flag is not
                                    # specified or is a space.
DataType = Streamflow Streamflow
                                    # Optional.  Data types for each time series
                                    # (consistent with TSID if specified).
                                    # The default is to use the data type in the TSID
                                    # Supplied to simplify use by other programs.
Units = CFS CFS                     # Optional.  Units for each time series
                                    # (default is no units).
MissingVal = -999 -999              # Optional. Missing data value for each
                                    # time series (default is -999).
                                    # NaN (not a number) is recommended
IncludeCount = true                 # Optional.  If true, column after date/time
                                    # is record count (1...) (default is false).
IncludeTotalTime = true             # Optional.  If true, column after date
                                    # is cumulative time (0...) (default is false).
# Both of above can be true, and both columns will be added after the date
Start = 1996-10-18:00:00            # Required.  Start date for time series
End = 1997-06-14:00:00              # Required.  End date for time series
                                    # Period dates should be of a precision consistent
                                    # with the dates used in the data section below.
# Optional.  The following line can be read into a spreadsheet or database for
# headers.  The lines above this line can be ignored in a spreadsheet import.
# The number of headings should agree with the number of columns.
Date "Time" "Count" "TotalTime" "Description 1" "DataFlag1" "Description 2"
1996-10-18 00:00 1 0 110.74 "m" 14.2
1996-10-18 00:15 2 15 113.24 "" 13.7
...
```

DateValue Files and Standard Time Series Properties

The standard time series identifier for DateValue files is as follows:

> Location.DataSource.DataType.Interval.Scenario~DateValue~PathToFile

Because DateValue time series files are a persistent storage format for in-memory time series objects, the properties stored in the file closely match the standard time series properties of the objects. In particular, the time series data type, units, and missing data value are consistent with time series header information. The TSID property in a DateValue file is directly applied to time series objects read from the file, allowing explicit identification of the time series in the file, regardless of the name of the file. This allows multiple time series to be saved in a single file. The data source typically agrees with that determined from a data-supplying agency or model that generates the data.

Limitations

DateValue files have the following limitations:

- The header information in DateValue files may be too technical for some general tools. However, simple delimited files require additional information for TSTool to properly handle the data as time series with properties. Spreadsheets can import DateValue files easily by ignoring the header lines.
- Because date/time values are included on every data line, processing DateValue time series files requires more disk space and processing time. However, using the dates on each line also allows gaps in data to be omitted from the file. Inclusion of the date/times for each data point is considered a reasonable trade-off to ensure data quality and readability. Many other time series file formats also include the date/time on each line.

Examples

The following is an example file for day interval data with data flags:

```
#
#
Delimiter   = " "
NumTS       = 1
TSID        = "MyLoc..MyData.Day"
Alias       = "MyLoc"
Description = "Test data, pattern"
DataType    = "MyData"
Units       = "CFS"
MissingVal  = -999.0000
DataFlags   = true
Start       = 1950-01-01
End         = 1951-03-12
#
# Time series comments/histories:
#
#
# Creation history for time series 1 (TSID=MyLoc..MyData.Day Alias=MyLoc):
#
#    Created new time series with interval determined from TSID "MyLoc..MyData.Day"
#    Set 1950-01-01 to 1951-03-12 to pattern=5.000,10.000,12.000,13.000,75.000,
#                       flags=Flag1,Flag2,,Flag4,Flag5
#
#EndHeader
Date "MyLoc, CFS" DataFlag
1950-01-01 5.0000 "Flag1"
1950-01-02 10.0000 "Flag2"
1950-01-03 12.0000 ""
```

```
1950-01-04 13.0000 "Flag4"
1950-01-05 75.0000 "Flag5"
1950-01-06 5.0000 "Flag1"
1950-01-07 10.0000 "Flag2"
1950-01-08 12.0000 ""
1950-01-09 13.0000 "Flag4"
1950-01-10 75.0000 "Flag5"
…
```

The following is an example file for hour interval data (no data flags):

```
#
#
Delimiter   = " "
NumTS       = 1
TSID        = "MyLoc..MyData.Hour"
Alias       = "MyLoc"
Description = "Test data, pattern"
DataType    = "MyData"
Units       = "CFS"
MissingVal  = -999.0000
Start       = 1950-01-01 00
End         = 1950-01-03 12
#
# Time series comments/histories:
#
#
# Creation history for time series 1 (TSID=MyLoc..MyData.Hour Alias=MyLoc):
#
#    Created new time series with interval determined from TSID "MyLoc..MyData.Hour"
#    Set 1950-01-01 00 to 1950-01-03 12 to pattern=5.000,10.000,12.000,13.000,75.000
#
#EndHeader
Date Time "MyLoc, CFS"
1950-01-01 00 5.0000
1950-01-01 01 10.0000
1950-01-01 02 12.0000
1950-01-01 03 13.0000
1950-01-01 04 75.0000
1950-01-01 05 5.0000
1950-01-01 06 10.0000
…
```