

Specification for Heliophysics Event List (Catalogue) Format ("HPEvent" format)

Version 1.0.4

Heliophysics Data Environment (HPDE)

January 31, 2020

Corrected March 19, 2022

Introduction

The Heliophysics Event List format is a plain text data format with metadata embedded in comments. The metadata is sufficient to properly interpret typical event lists and can point to other sources of information related to the data. The format is designed to be easily parsed and to be easily converted to a VOTable format (see below).

Definition of the Heliophysics Event List Format

The Heliophysics Event List Format (HPEvent) is a text data format with header information. The Heliophysics Event List is defined as follows:

1. A file or stream of text containing one or more lines.
2. A line is a sequence of characters ending with a new line (<LF>) character. A line may also end in a <CR><LF> sequence, in which case the <CR> is ignored.
3. A line beginning with a number sign (#) is a header line. All other lines are data.
4. A header line which begins with a single word followed by a colon (:) and space contains tagged metadata. The word is the tag name (keyword). The value is the text following the ":" and space or the entire line if a keyword and colon is not present.
5. Header lines can only occur before the data.
6. Data is represented as fields delimited by one or more spaces.
7. Double quotes (") are used around any text entries containing spaces.

Representation of Time

All time values (both metadata and data) are represented as an ISO 8601 Date and Time [1]. Values are given in the format (yyyy-mm-ddThh:mm:ss.sssZ). The time value of "0001-01-01T00:00:00.000Z" is considered a flag for an unknown or unspecified time value.

Reserved Keywords

Required

EventTableVersion: The specification name, version and URL to a detailed description of the format specification. The format of the value is: Name Version URL.

ListTitle: Title of the Event List.

FieldNames: Ordered, space separated names of fields in the data.

FieldUnits: Ordered, space separated units for each field.

FieldTypes: Ordered, space separated types for each field.

FieldNulls: Ordered, space separated fill values for each field.

Optional (recommended)

ListID: SPASE ID of list if available.

CreationDate: ISO 8601 Date and Time of the Table creation.

ModifyDate: ISO 8601 Date and Time of the last modification of the list.

ListBeginDate: First possible ISO 8601 Date and Time for the List.

ListEndDate: Last possible ISO 8601 Date and Time for the List.

Contact: Relevant contact person and possibly contact information.

ContactID: SPASE ID of contact person, if available.

Description: Free text which continues until the next keyword. It may have multiple lines. Each line is formatted as a comment and must begin with "#".

Metadata requirements

The first line must be a header line with the EventTableVersion metadata.

Header lines with ListTitle, FieldNames, FieldUnits, and FieldTypes metadata required.

For FieldNames, FieldUnits, FieldTypes and FieldNulls the number of values in the ordered list must match the number of fields in the data. Values are delimited with one or more spaces.

The first variable listed in FieldNames must be the field name for the beginning times of the events, and the second variable must be the field name for the end times. For an event which occurs in an instant of time (a single point) the beginning and end time for the event should be the same.

If additional date/time fields exist, it is recommended that they are listed after the end time field and before all other available fields.

Every field must be assigned a name.

FieldNames may not contain spaces or special characters except "_" (underscore).

FieldUnits may not contain spaces, but may contain "*", ".", "/", "^", "-", and "+". The unit for ISO 8601 date/times must be "DateTime". The units for generic text fields must be "char". The value for a field without units is "None".

Possible FieldTypes are: boolean, short, int, long, char, float, double, floatComplex, doubleComplex; These are identical to those used by VOTable [2], other VOTable types are not allowed, i.e., bit, unsignedByte, and unicodeChar. A date/time field has a field type of "char".

The ListBeginDate and ListEndDate keywords give the start and stop times of the surveyed period. All events must fall within this range. The time flag value "0001-01-01T00:00:00.000Z" should be used for time values that are not known or not relevant. If both ListBeginDate and ListEndDate values are given as the time flag value, the list is to be considered a non-exhaustive list of events.

It is recommended (but not required) that the metadata block finishes with a line containing properly ordered space-separated field names (FieldName values), preceded by #.

Data Requirements

All fields should be filled either with data entries or with fill values of the same type as the quantities.

If the events are considered to be instantaneous, then the end times should be entered as the same as the beginning times to avoid confusion with other times in the table.

Further Notes

A description may (should!) include the way the HPEvent list has been generated or its origin if it has been adapted from an existing referenced HPEvent list. URLs to additional information are desirable.

The observatories and variables used to create the table should be identified and described well enough to be understood. If a SPASE ID exists for a referenced item it should be included in the description.

Any table based on a well-developed procedure should provide an information URL on the procedure and a contact person or group. Once registered, a SPASE ID should be given as the event list name.

There are no restrictions or standards for the contents of the columns of the table after the second.

VOTable Version

The VOTable version of the HPEvent formatted lists has all the same metadata but in a different form.

The tags that begin with "Field" are mapped to VOTable Fields, with each variable in the list corresponding to a separate field. The beginning and end times of the events must be the first two fields. If there are other date/time fields, they should be listed next.

The general syntax is:

```
<FIELD datatype="<FieldType>" unit="<FieldUnit>" name="<FieldName>">
  <VALUES null="<FieldNull>"/>
</FIELD>
```

where the FieldName, FieldType, etc. correspond sequentially to the individual fields in the ASCII table lines.

Fields that have units of "char" must also have an arraysize:

```
<FIELD arraysize="*" datatype="char" unit="text" name="<FieldName>">
  <VALUES null="<FieldNull>"/>
</FIELD>
```

and those with a unit of "DateTime" need a ucd and xtype:

```
<FIELD arraysize="*" datatype="char" xtype="DateTime" name="<FieldName>" ucd="time.epoch" >
  <VALUES null="<FieldNull>"/>
</FIELD>
```

The ListTitle maps to:

```
<RESOURCE name="<ListName>" >
```

with </RESOURCE> closing the block at the end of the actual table entries.

All other tags in the ASCII table version are mapped to VOTable Parameters, given by <PARAM ... />. The ListStartDate and ListStopData are combined within a Group (<GROUP />), as are Contact and ContactID.

The actual text of the Description is contained in a CDATA block:

```
<DESCRIPTION>
<![CDATA[
  Test of the description goes here with the same content
  as in the ASCII table version.
]]>
</DESCRIPTION>
```

The column entries in the flat ASCII version of the HPEvent lists are replaced by tagged tables, with each row delimited by <TR> ... </TR> and the row entries by <TD> ... </TD>.

Implementation Notes

ABNF[3] grammar (basic format)

```
file = 1* header 1* record
header = # [keyword : SP] text [CR]LF
record = field *( 1* SP field) [CR]LF      ; SP is space
field = ["] 1* (VCHAR / SP) ["]           ; All visible characters and space
```

Parsing data

Examples of parsing space+ delimited data records with quoted values.

Perl

```
use Text::ParseWords;
@fields = quotewords('\s+', 0, record);
```

Note: Text::ParseWords properly handles quoted values.

Java

```
String[] fields = record.split("[^\\s\\\"]+|\"([^\"]*)\"");
```

Note: Properly handles quoted values.

References

- [1] ISO 8601:2004, http://dotat.at/tmp/ISO_8601-2004_E.pdf
- [2] VOTable Format Definition, Version 1.2, IVOA Recommendation 2009-11-30 (Section 6, "Definitions of Primitive Datatypes ") <http://www.ivoa.net/Documents/VOTable/20091130/REC-VOTable-1.2.html>
- [3] Crocker, D. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", RFC 2234, November 1997.

Appendix 1: Example of a Plain ASCII HPE Data File

Sector event boundary event list:

```
# EventTableVersion: HPEevent 1.0 http://spase-group.org/docs/conventions/HDMC-Event-List-Specification-v1.0.pdf
# ListTitle: Svalgaard_Heliospheric_Sector_List
# ListID: spase://VSPO/Catalog/Svalgaard_SB_List
# CreationDate: 0001-01-01T00:00:00.000Z
# ModifyDate: 0001-01-01T00:00:00.000Z
# FieldNames: time_begin time_end plus_to_minus daysbefore daysafter
# FieldUnits: DateTime DateTime None days days
# FieldTypes: char char boolean int int
# FieldNulls: 0001-01-01T00:00:00.000Z 0001-01-01T00:00:00.000Z 0 0 0
# ListBeginDate: 1926-01-01T00:00:00.000Z
# ListEndDate: 2012-01-01T00:00:00.000Z
# Contact: Leif Svalgaard
# ContactID: spase://SMWG/Person/Leif.Svalgaard
# Description:
# IMF Sector Boundaries 1926-2011, compiled by Leif Svalgaard
# This table identifies days in the solar wind upstream of the Earth when the
# polarity of the interplanetary magnetic field changes, indicating that the
# field now connects to a region of opposite magnetic polarity on the Sun. At
# solar minimum this would typically mean connecting to the coronal hole at the
# opposite pole of the Sun.
#
# Note that the original '+,-' column has been replaced with a "plus_to_minus"
# variable that is 1 if the sector goes from + to - and 0 for - to +
#
# A description and the original list can be found at: http://wso.stanford.edu/SB/SB.html
# The documents Svalgaard references below can be found at:
# http://www.leif.org/research/, Entry 20.
#
# Description from Leif Svalgaard:
#
```

```

# The last two numbers are the number of
# days of each polarity on each side of the boundary, so "+,- 2010 01 06
# 16 4" means that there are 16 '+' days before the boundary and 4 '-'
# days after. This is to make it easy to do analysis, e.g. isolated SBs.
#
# The method of deciding is basically the same as in the earlier Svalgaard
# atlases: Atlas of IMF Polarity 1947-1975.pdf (SUIPRs 629 and 648) also
# found on my website. Basically I make the 'best' guess, IMF having
# double weight of ground-based. This makes the list non-homogeneous, but
# so be it. Four days of duration is my general criterion for a sector,
# but I'm not religious about it, and occasionally, thinner sectors are
# included if they are clear and recurrent. Using the numbers on the right
# you can make your own selection.
#
# In the past, I did the assignment 'by eye', now I have a simple computer
# algorithm to do it. It involves a slight smoothing to essentially
# disregard mixed or too tiny sectors. No visual quality control has been
# performed [I trust my algorithm more than my personal bias].
#
# [It's a good idea to repeat the variable names right above the table]
# time_begin          time_end    plus_to_minus  daysbefore  daysafter
1926-01-29T00:00:00.000Z 1926-01-30T00:00:00.000Z    0 21  5
1926-02-10T00:00:00.000Z 1926-02-11T00:00:00.000Z    1  5 15
1926-02-25T00:00:00.000Z 1926-02-26T00:00:00.000Z    0 15  5
1926-03-02T00:00:00.000Z 1926-03-03T00:00:00.000Z    1  5 12
1926-03-22T00:00:00.000Z 1926-03-23T00:00:00.000Z    0  5  6
1926-04-06T00:00:00.000Z 1926-04-07T00:00:00.000Z    1  6  6
1926-05-04T00:00:00.000Z 1926-05-05T00:00:00.000Z    1  8 13
1926-05-17T00:00:00.000Z 1926-05-18T00:00:00.000Z    0 13 15
1926-06-01T00:00:00.000Z 1926-06-02T00:00:00.000Z    1 15  4
1926-06-15T00:00:00.000Z 1926-06-16T00:00:00.000Z    0  9 10
1926-06-25T00:00:00.000Z 1926-06-26T00:00:00.000Z    1 10  6
1926-07-12T00:00:00.000Z 1926-07-13T00:00:00.000Z    0  9 19

```

1926-07-31T00:00:00.000Z	1926-08-01T00:00:00.000Z	1	19	8
1926-08-08T00:00:00.000Z	1926-08-09T00:00:00.000Z	0	8	11
1926-08-19T00:00:00.000Z	1926-08-20T00:00:00.000Z	1	11	5
1926-09-07T00:00:00.000Z	1926-09-08T00:00:00.000Z	0	11	6
1926-09-13T00:00:00.000Z	1926-09-14T00:00:00.000Z	1	6	4
1926-09-17T00:00:00.000Z	1926-09-18T00:00:00.000Z	0	4	9
1926-09-26T00:00:00.000Z	1926-09-27T00:00:00.000Z	1	9	11
1926-10-16T00:00:00.000Z	1926-10-17T00:00:00.000Z	0	7	13
1926-11-05T00:00:00.000Z	1926-11-06T00:00:00.000Z	1	4	17
1926-12-04T00:00:00.000Z	1926-12-05T00:00:00.000Z	1	6	8
1926-12-12T00:00:00.000Z	1926-12-13T00:00:00.000Z	0	8	19
1927-02-12T00:00:00.000Z	1927-02-13T00:00:00.000Z	0	4	12
1927-02-24T00:00:00.000Z	1927-02-25T00:00:00.000Z	1	12	6
1927-03-10T00:00:00.000Z	1927-03-11T00:00:00.000Z	0	5	6
1927-03-25T00:00:00.000Z	1927-03-26T00:00:00.000Z	1	7	14
1927-04-08T00:00:00.000Z	1927-04-09T00:00:00.000Z	0	14	15
1927-04-23T00:00:00.000Z	1927-04-24T00:00:00.000Z	1	15	4
1927-05-05T00:00:00.000Z	1927-05-06T00:00:00.000Z	0	6	14
1927-05-19T00:00:00.000Z	1927-05-20T00:00:00.000Z	1	14	6
1927-06-03T00:00:00.000Z	1927-06-04T00:00:00.000Z	0	6	12
1927-06-15T00:00:00.000Z	1927-06-16T00:00:00.000Z	1	12	16
1927-07-01T00:00:00.000Z	1927-07-02T00:00:00.000Z	0	16	16
1927-07-17T00:00:00.000Z	1927-07-18T00:00:00.000Z	1	16	12
1927-07-29T00:00:00.000Z	1927-07-30T00:00:00.000Z	0	12	7
1927-08-13T00:00:00.000Z	1927-08-14T00:00:00.000Z	1	5	17

... Truncated for brevity

Change history

2023/02/03

Chuck Dolan

Pushed 2022/03/19 update to server