

tsp Ground Temperature Standard

this is version 1.0 of the tsp ground temperature data standard. It was designed to reduce ambiguity, formalize many conventional patterns in ground temperature data files, and offer flexibility to handle data and metadata in a single, human-readable, and easily editable file. It can be extended to support a wider variety of permafrost data such as interval data and other time-series data (see details below).

Requirements

1. File requirements:

1. The file name should use the `csv` extension
2. The file must use UTF-8 (without BOM) encoding
3. The comma must be used as a separator

2. Metadata lines:

1. The file may begin with any number of lines beginning with `#`. These can contain site metadata or other information.
2. Metadata key-value pairs should be separated by an equals sign (`=`)
3. Spaces before and after metadata keys/values are allowed for readability but should be ignored (i.e. both `#longitude=-114.23` and `# longitude = -114.23` are valid and equivalent).
4. Keys should be lowercase with underscores (e.g. `sensor_type` not `Sensor Type`)
5. Metadata should use the standardised keys (Section 8)

3. Header line:

1. The file must contain a single header line to define data columns
2. The header line must be the first line not beginning with `#`
3. The first element of the header line must be 'timestamp'
4. Header values must not be blank; the header line must not have a trailing comma

4. Data structure: the file must use one of two data formats:

1. **Wide format (depth as columns)** *Note: Wide format has no provision for multiple sites in one file. If this is needed, use long format*
 1. The remaining elements of the header line must be depth values, measured in metres, using exclusively numerical characters [0-9] with an optional decimal (.) or negative sign (-).

2. Depth values must be in ascending order (shallower to deeper)
3. Depth values must be unique
4. Timestamps must be in chronological order
5. Timestamps must be unique

2. Long format (depth as data)

1. The header line must contain a 'depth' column
2. The header line must include a 'temperature' column
3. The file may have any number of additional unique column names as elements in the header line
4. Additional header names must only use lowercase letters (a-z), numbers (0-9), and underscores (_), and must start with a letter.
5. The header line may include a field `site_id`
6. If a `site_id` column is not present, timestamp-depth combinations must be unique across rows,
7. If a `site_id` column is present, timestamp-depth-site_id combinations must be unique across rows
8. Depth values must not be missing (empty strings not allowed in depth column)
9. Temperature values may be missing (indicated by empty string)
10. If a `site_id` column is not present, timestamps must be in chronological order
11. If a `site_id` column is present, rows should be grouped by site_id, then either: (a) by depth with timestamps in chronological order within each depth, or (b) by timestamp with depths in ascending order within each timestamp

5. Time representation:

1. Timestamps must use ISO 8601 extended format (`YYYY-MM-DD` or `YYYY-MM-DDTHH:MM:SS`)(ISO 8601-1:2019). Fractional seconds are permitted.
2. Timestamps should include time zone information when available (e.g. `2002-04-03T10:30:00-06:00` , `2022-01-07T11:30:00Z`)
3. When time zone is unknown, timestamps without timezone are permitted (e.g. `2002-04-03T10:30:00`) and are assumed to be local time
4. Timestamps with timezone information should preferably use UTC (Z)
5. Timestamp values must not be missing

6. Units and precision:

1. Temperatures must be in units of degrees C
2. Depth values must be recorded in metres
3. Depth values must be oriented positive downward (0.1 is shallow, 100 is deep, -1 is above the ground surface)

7. Missing data:

1. Missing data must be indicated by an empty string (e.g. ,). Placeholder values (e.g. NA, NaN, -999, NULL, or any other placeholder) are not allowed.
2. Trailing commas are permitted in data rows and represent missing data in the last column
3. The file may not have any blank lines, with the exception of the final line.

8. **Standardized metadata keys:** The following metadata keys, units, and definitions should be preferred over ad-hoc formulations:

- latitude: the latitude, measured in decimal degrees, at the measurement location.
- longitude: the longitude, measured in decimal degrees, at the measurement location.
- elevation: the ground surface elevation, in metres, at the measurement location.
- site_id: a unique name or identifier for the measurement location.

Examples

Example for wide format

```
# site_id = whitepine lake
timestamp,0.1,3.5,4.25
2025-01-01T13:00:00Z,1.2,-1.6,-2.4
2025-01-01T14:00:00Z,, -1.6,-2.4
2025-01-01T15:00:00Z,1.6,-1.6,-2.4
```

Example for long format

```
# latitude = 62.6421
# longitude = -114.3231
# elevation = 420
timestamp,depth,temperature,quality_flag
2025-01-01T13:00:00-06:00,1,4.2,
2025-01-01T13:00:00-06:00,2,-0.1,
2025-01-01T13:00:00-06:00,4,-2.3,4
2025-01-02T13:00:00-06:00,1,1.5,2
2025-01-02T13:00:00-06:00,2,,1
2025-01-02T13:00:00-06:00,4,-2.31,
```

Extended format for additional use cases:

The tsp format may be extended for additional use cases relevant for permafrost (interval measurements and non-temperature time series).

Long format (Extended)

1. The 'depth' column may be replaced by two columns: 'depth_from' and 'depth_to'
2. For every row, the 'depth_from' value must be less or equal to the 'depth_to' value (when depth_from equals depth_to, this represents a point measurement at that depth)
3. For a given timestamp (and site_id if present), depth intervals must not overlap
4. The 'temperature' column may be omitted if other measurement columns are present
5. When temperature is omitted, at least one other measurement column must be present

Examples

Interval data

```
# latitude = 62.6421
# longitude = -114.3231
# elevation = 420
timestamp,depth_from,depth_to,visible_ice
2025-07-01,0.0,1.2,0.00
2025-07-01,1.2,1.5,0.10
2025-07-01,1.5,2.5,0.15
2025-07-01,2.5,2.8,0.28
2025-07-01,2.8,2.9,0.11
```

Non-temperature sensors

```
# latitude = 62.6421
# longitude = -114.3231
# elevation = 420
timestamp,depth,soil_moisture
2025-01-01T13:00:00-06:00,1,0.30
2025-01-01T13:00:00-06:00,2,0.30
2025-01-01T13:00:00-06:00,4,0.32
2025-01-02T13:00:00-06:00,1,0.33
2025-01-02T13:00:00-06:00,2,0.32
2025-01-02T13:00:00-06:00,4,0.34
```