ISD Format Specification and Conventions

Language Specification ver 1.0

Wednesday, March 12, 2014

1. Introduction

This document defines the specification of ISD format, a binary format to describe version 1.

2. Summary

The following is the summary of the fields defined in ISD format.

header

>*prologue*

>>*magic (4byte)*

>>*version (1byte)*

>>*endian (1byte)*

>>*padding (2byte)*

>>*timestamp (20byte)*

>>*num_variables (uint32_t)*

>>*num_bytes_comment (unit32_t)*

>>*num_bytes_descs (unit32_t)*

>>*num_bytes_units (unit32_t)*

>*comment*

>*descriptions*

>>*+description*

>*units*

>>*+unit*

body

where the sections with '+' label can be multiple.

3. ISD Format

As a notation, the field name of the format are surrounded with '*'.

And ISD format file contains a *header* immediately followed by a *body*.

The *header* starts with a fixed-length *prologue* part immediately followed by *comment* part of arbitrary length, then a variable-length *descriptions*, and ends with optional variable-length *units*.

The *prologue* part consists of *magic*, *version*, *endian*, *padding*,*timestamp*, *num_variables*, *num_bytes_comment*, *num_bytes_descs*, and *num_bytes_units* in the respective order.
The *magic* is 4 byte-long consecutive block, which always filled with "ISDF".
The *version* is 1 byte (char) filled with the ISDF format version, currently 1.

The *endian* is 1 byte (char) designating the byte order for data words in this file. It is filled with 1 if the endianness is little-endian, or with 0 otherwise i.e. big-endian.

The *padding* is 2 byte-long consecutive block, which filled with 0 (NULL).

The *timestamp* is a 20 byte-long string representing a timestamp in the date and time format defined in [RFC3339] with "Z" as its time-offset. For example, "1990-12-31T23:59:60Z" is a valid string for *timestamp*.

The *num_variables* is of type uint32_t, which value is the number of variables.

The *num_bytes_comment* is of type uint32_t, which value is the byte-length of the following *comment* in the *header*.

The *num_bytes_descs* is of type uint32_t, which value is the byte-length of the following *descriptions* in the *header*.

The *num_bytes_units* is of type uint32_t, which value is 0 if *units* is omitted, or the byte-length of the following *units* in the *header* otherwise.

The *comment* is an array of bytes.

The *descriptions* consist of multiple *description*s; each of them corresponds to a single variable, so there are as many *description*s as the number of variables. The first 4 bytes (uint32_t) of *description* specifies the byte-length of the *title* of the corresponding variable, and then the string of the *title* follows.

The *units* consist of multiple *unit*s; each of them corresponds to a single variable, so there are as many *unit*s as the number of variables. If the first 4 bytes (uint32_t) of a *unit* has value 0, then its corresponding variable's unit is omitted. Otherwise the value specifies the byte-length of the unit name of the corresponding variable, and then the string of the unit name, e.g. "millisecond", follows.

The *body* consists of zero or more *step*s; each of them corresponds to values of all of variables at a time step.

Every *step* has the same, fixed length of (the number of variables) * 8 (double) bytes.

Each double value in the step corresponds to the value of a respective variable in the order specified in the *descriptions*.

References

[RFC3339] http://tools.ietf.org/html/rfc3339

4. ISD convention

This document explains the convention for better interoperability between software applications exchanging files in the ISD format.

4.1. On title of variables

Each title of variables in an ISDF file should be unique among them.

The following ABNF in the flavor of [RFC5234] defines the syntax for a variable's title formally:

```
title = labeled / unlabeled
```

labeled = unlabeled "@" label

label = 1*VCHAR

unlabeled = prefixed / name

prefixed = UUID ":" name

name = ALPHA *("_" / ALPHA / DIGIT)

where

UUID is as defined in [RFC4122].

4.2. On Timeseries data

If an ISDF file contains the variable representing the time, then name it "*time*" and make it the first variable in the data.

References

[RFC4122] http://tools.ietf.org/html/rfc4122
[RFC5234] http://tools.ietf.org/html/rfc5234