Definition of the standard data format for skyglow observations

Version 0.1.5
July 7, 2012

This document describes Version 0.1.5 of the community standard for measurements of skyglow. In each section of this document the standard is described, and in many cases the rationale behind the decision to structure the file in this way is included. Two example files are provided in order to show what the standard looks like in practice. It is the expectation of the skyglow monitoring and research community that any newly developed system for measuring skyglow will follow this same data format.

Character set, line breaks, and line length

- All character encoding is to be done using the ASCII format. Although this precludes the use of non-western characters, it ensures that data files will be universally readable.
- The expected line break sequence is "\r\n". This ensures that data from different platforms will look the same.
- Tabbing must not be used anywhere in the data file.
- With the exception of the "user comments" fields, header lines must not exceed 150 characters in length.
- No blank lines are allowed to occur anywhere in the file
- Trailing zeros must be used in decimals where the precision is specified (e.g. 10.50 not 10.5)

Header

Each header entry begins with a leading # followed by a space separator (this is in order to allow easy filtering of the header when reading in data). While the data format is designed to make transfer of data easier, the header is designed to be human readable. For this reason, comma separation is used in cases where the header entry has more than one variable. The maximum length of a header line is not fixed, but we strongly recommend that a limit of 150 characters be observed. The values stored in the header are as follows:

Data format version

The current version is 0.1.5. This will be updated to 1.0 when the format is finalized.

Document type definition

Provides a link to the definition of the format. The current format will be hosted at http://www.darksky.org/LANstandards/clan_format_0.1.5.pdf.

Number of header lines

The total number of lines in the header is listed for two reasons. First, to allow data analysis programs to skip the header, and second to allow those programs to check that the header was complete.

Data license information

We encourage everyone measuring skyglow to release their data under the Open Database Licence (OdbL, http://opendatacommons.org/licenses/odbl/). We understand that some data providers may have a need to temporarily embargo their initial data before joining a present or future global network, and for this reason it is a good idea to record in the data file itself whether the data may be shared or not.

Type of device

Examples: SQM-LE (and -LU, -LU-DU, etc), IYAL, NSBM, DigiLum

Instrument ID

The instrument ID is a human readable unique name. Since the number of stations doing monitoring is currently small, it probably won't be a problem for users to come up with a name on their own. In the future, when a skyglow measurement database is established, there should be a procedure to have names assigned. In the meantime, please register your name with christopher.kyba@wew.fu-berlin.de. Spaces must be avoided in Instrument IDs, please use dash or underscore instead. Example instrument names: SQM-RIVM1, Dahlem_tower_le

Data supplier

The institution and/or person who was responsible for setting up and acquiring the data. Since it is expected that skyglow data will be archived for generations, detailed contact information (e.g. email address) is unlikely to be as helpful as information about the institute that supplied the data. Users are free to provide contact information in the user comments section below.

Location name

A unique, human readable name describing the geographic location. In the future, names for permanent stations must be proposed to whoever is operating the skyglow database. For now please register your name with christopher.kyba@wew.fu-berlin.de. Example location names: Berlin-Dahlem, Schipluiden

Location co-ordinates

The location at which the device is located (for permanent stations), or alternatively the starting location for a moving station. The co-ordinates are given in decimal degrees in the order latitude, longitude, elevation. Latitude is defined to run from -90 to +90, longitude from -180 to +180, and each is to be given to four decimal places. Elevation is to be reported to the meter, with no decimal. No character is to be used to describe the hemisphere (negative values represent locations in the Western or Southern hemispheres). Station elevation is given in meters above nominal sea level according to the WGS 84 datum (this is what you get from the current GPS satellites).

Local timezone

The local time is recorded because some factors contributing to skyglow depend upon local rather than celestial time (e.g. sports field lighting turning off at 22:00). Because local time is managed so differently in different parts of the world and can change due to the passage of new laws, it is dangerous to ask users to define the local time themselves. Furthermore, defining the local time relative to UTC in the header leads to problems on the nights when daylight savings time is switched on and off. Luckily, there exists a software solution in the form of the IANA timezone database.

This database keeps track of the time zone information for all locations worldwide, and many users have likely already unknowingly interacted with it when they chose their timezone when setting up their computer. If users specify a locale from the IANA timezone database then there will never be a question as to the time zone, and political changes to the timing of summer time will be automatically handled. In order to make this possible, **it is essential that you select a timezone from the tz database**. You can find your timezone at one of these links:

- http://php.net/manual/en/timezones.php
- http://en.wikipedia.org/wiki/List_of_tz_database_time_zones

Examples of allowed timezones:

- America/Mexico_City
- Asia/Hong_Kong
- Europe/Berlin

Examples of incorrect timezones:

- ~~USA/Mountain Standard Time~~
- ~~Central European Summer Time~~
- ~~Europe/Hogwarts~~

Time synchronization

For long time series the recorded time ought to be synchronized to a global clock. For many locations this will be done automatically through the network time protocol (NTP). Remote stations may make use of GPS, or may have no time synchronization at all (e.g. SQM-LU-DU).

Examples: NTP, GPS, Approximately monthly by hand, free running – synched by GPS at 2012-05-30T15:30:05 local time

Moving or stationary observatory

A record of whether the observing platform is STATIONARY or MOVING. The data format may be slightly different for the two, because moving platforms may also include locations with the data (if this is supported by the device). See also the section on "**moving platforms**" below.

Scanning or fixed look direction

A record of whether the look direction is FIXED or SCANNING. The data format is slightly different for the two, because scanning platforms must report the look direction with each measurement. See also the section on "**scanning platforms**" below.

Number of channels

Some devices, such as the Night Sky Brightness Monitor, have more than one channel that could look in different directions, or perhaps through different filters. The standard was designed to be flexible in these cases.

Filter used for each channel

This field is included to anticipate the potential use of non-standard filters for any given device. If a custom filter is used then details about it are to be given in the comments section. If the data goes into a global database then a table with filter parameters will need to be included. Note that since there is no restriction on the length of a header line, it would be possible to include the entire filter curve as a user comment. If this is done we recommend the repeating format "wavelength1, transmission1, wavelength2, transmission2, etc." For a new device for which a sensitivity curve has not yet been published, it would be possible to put the transmission curve after the comments. (Of course, if you've built a custom device and measured the transmission and/or sensitivity curves, then you really ought to write up a description of what you've done!)

Examples: HOYA CM-500 (standard SQM filter), Johnson V, None

Measurement direction

This is the direction in which each channel of the device "looks", given in decimal degrees of zenith and azimuthal angles and for each channel (in that order). Each entry is separated with commas. Zenith angle runs from 0^o at zenith to 90^o at the horizon, and 180^o for directly down. Azimuthal angle increases clockwise from 0^o at directly North. For scanning devices, record the initial look direction at the time the header is written. Because we can't anticipate the pointing precision of future devices we don't propose a required degree of reporting precision in the look direction.

Examples:

Zenith look direction is 0, 0
Northern horizon look direction is 90, 0
Eastern horizon look direction is 90, 90
Southern horizon look direction is 90, 180
Western horizon look direction is 90, 270

Example for a two channel device: 0, 0, 30, 245

Fields of view for each channel

This is mainly intended to provide a minimum of data for custom devices. If the data from such devices is included in a global database, then detailed information will need to be made available through the database (for example, by including a PDF document describing the device). If the field of view doesn't have a sharp cutoff then the full width at half maximum is to be used. Like all angles, field of view is reported in degrees.

Some devices may have a two dimensional field of view (e.g. cameras). In this case the two angles are to be specified separated by a comma, and the user must describe the orientation in the header entries specific to the device. Note that this format is only intended for devices with a small amount of data. (In the case of a camera's pixels being summed, it would be possible to completely avoid this problem by only selecting the pixels enclosed up to a certain zenith angle, giving a perfectly defined FOV).

Number of fields in each data line

This will be 12 (for the universal and local time) plus the number of data entries that come from the device.

Device specific characteristics

Several header lines will be needed to describe individual devices. These lines will include for example the firmware version and internal calibration parameters. For example, for the SQM these are:

- **SQM serial number** – you can get this from the “ix” command.
- **SQM firmware version** – you can get this from “ix”. (SQM-LE users: be sure you are running at least version 4-3-21.)
- **SQM cover offset value** – this is the difference in reading caused by the weatherproof housing. For people using the standard housing from Unihedron the value is -0.11.
- **SQM readout test ix -** these tests allow for a check that the SQM is reading out data correctly
- **SQM readout test rx** – these tests allow for a check that the SQM is reading out data correctly
- **SQM readout test cx** – this reports the internal calibration parameters of the SQM

User comments

Five lines are reserved for user comments. Users can include any information that they wish, for example their contact information or details about the surroundings that might be useful to someone who analyzes the data. These lines are not subject to the 150 character limit, but note that a database may cut off lines that run longer than this.

Extra blank lines

This allows for future changes to the header that make it longer. (Of course, careful programmers will check the number of header lines before they skip it, as well as checking the version number and end of header statement to be sure that the header is as expected.)

Data line entries in human format

This line describes the data fields that follow the header in human readable terms. See example files.

Units of data line entries

This line indicates the units in which each datum is recorded (e.g. hours, Hz, degrees C). Unlike the rest of the header, here each word is separated with the data delimiter rather than a comma.

End of header statement

“# END OF HEADER” indicates that the header was written completely. All entries that follow below are data. Remember that blank lines must never occur.

Data lines

All recorded data must be the raw data reported from the device. This file format is meant to record "level0" data that has not been processed in any way, and will never be replaced after a reprocessing. This means, for example, that the "glass cover" correction for SQM data must not be applied, and that values reported by the IYA Lightmeter are to be stored as counts, not converted by software to lux. The only exception to this rule is temperature, which is always reported in degrees Celsius.

All data lines begin with a universal and local time stamp. UTC time was chosen over other universal options (e.g. JD, unix time) due to its improved human readability and ease in handling leap seconds.

Dates and times

Dates and times are given according to ISO 8601, where the only allowed literal format is YYYY-MM-DDThh:mm:ss.fff. There must be no time zone tag present and times are given in 24 hour format. The first time in each line is UTC, the second time is the time in the local timezone. An example line for a hypothetical device located in the Pacific/Auckland timezone is shown below. UTC time is in red, local time (NZST) is in blue, data is in black. Note that the date differs between UTC and New Zealand standard time.

2012-07-07T13:03:32.000;2012-07-08T01:03:32.000;19.45

Data delimiter

The data delimiter is not yet fixed and is currently being voted on. The most popular candidates were bar, comma, semicolon, and space (i.e. "|" "," ";" and " ").

Missing or n/a data

If data is not applicable for some reason, then a value of "n/a" (always lower case) is to be written out. If data is corrupted, or not provided by the device, the field is left empty. This rule applies to both the header and the data.

Weather and other data

Many lightmeter operators are also operating weather stations. This is excellent, and we hope and expect that a future database will be able to also take in such data. But it was decided at the CLIC workshop that for reasons of simplicity and compatibility, weather data must not be included in the SQM data stream. All users are of course free to write out two data files, one according to this standard and another for backwards compatibility with their own analysis programs.

Similarly, if an lightmeter is operated inside of an observatory with a retractable roof, the user is responsible for ensuring that data is only written to the file when the roof is open. While this means some extra programming work for these users, it is much more likely that the person operating the observatory will be able to figure out how to process the data stream properly (i.e. rejecting roof-closed data) than that a remote person operating a database will be able to figure out when your roof was open.

Moving platforms

In the case of a device which writes out both the measured value and a location, five additional entries are to be included in the data stream. These are: latitude, longitude, elevation, speed, satellites.

Latitude, longitude, and elevation are recorded as specified in "**location co-ordinates**" above, speed is recorded in m/s, and satellites is the number of satellites used in the GPS fix. Latitude and Longitude are given to four decimal places, elevation is to be given in meters with one decimal place.

Scanning platforms

In the case of a device which does not have a consistent look direction (e.g. a scanning device or an SQM mounted on a telescope), two additional entries are to be included for each channel. These are the look direction in zenith and azimuth angles as described in "**Measurement direction**" above. A hypothetical example for a device with two channels, one of which is scanning from zenith to the Eastern horizon, is below. Data for Channel 1 is shown in blue, channel 2 in red:

2012;05;11;00;00;04.000;2012;05;11;00;00;04.000;1021;1021;0;0;0;90
2012;05;11;00;00;09.000;2012;05;11;00;00;09.000;1022;1057;0;0;15;90
2012;05;11;00;00;14.000;2012;05;11;00;00;14.000;1021;1170;0;0;30;90
2012;05;11;00;00;19.000;2012;05;11;00;00;19.000;1021;1280;0;0;45;90
2012;05;11;00;00;19.000;2012;05;11;00;00;19.000;1024;1565;0;0;60;90
2012;05;11;00;00;24.000;2012;05;11;00;00;24.000;1019;2129;0;0;75;90
2012;05;11;00;00;29.000;2012;05;11;00;00;29.000;1021;8956;0;0;90;90

File names, and frequency of starting new files

The format of the filename is: YYYYMMDD_HHmmss_[Instrument ID].dat

Example: 20120427_120000_SQM-RIVM1.dat

A new file should be started each day, at a time chosen by the user.

Files must contain only one header. If for some reason the data taking is interrupted and must be restarted, and if there is any doubt as to whether the time stamp, calibration, or other header parameters have changed, then a new file should be opened and a new header written. Data should be written in this new file until the standard "new file start time" of the user.

Readout frequency

Stationary stations should make observations in the range of once every 1-10 seconds. The absolute minimum readout is once per minute. (This requirement is waived for stand alone devices that run on battery power, they should read out as frequently as is reasonably possible.)

Requirements for future database

It was decided very early on in the development of this format that flagging of "bad" data is best done at a later stage, and not in the raw (level0) data itself. This is because it will often be the case that an operator learns of a problem only after the data has been acquired. For this reason, it is essential that a future database of light pollution measurements allow lightmeter operators the possibility to flag certain periods of time as not being fit for analysis. Other notable events (such as dates at which the device window was cleaned) should also be accommodated by the database.

Notes

A preliminary version of this standard was defined at the Cabauw Lightmeter InterComparison Workshop in May, 2012. It was edited to incorporate feedback from light pollution observers and researchers worldwide. Thanks to everyone who provided their valuable suggestions and constructive criticism!