INSTRUCTIONS FOR

EDDY COVARIANCE TURBULENT FLUX RAW DATA FORMAT

Version	Release date	Summary of changes
20171013	20171013	Changed file name structure requiring 2 digits for logger and file ID. Clarified how to submit TIMESTAMP and missing values. Clarified the header format
20171020	20171020	Corrected an error in the units of the GA Diagnostic
20180419	20180419	Removed from the list of the optional variables (Table 1B) SOS and FLOW_PRESSDROP because can not be collected.
20180810	20180810	Added a special case when an external pump is used. Added in Table 1a and 1b the correspondence between ICOS labels and LI-COR software. Clarified that the non-ICOS variables are stored but not processed. Other minor edits

The ICOS protocols and the derived Instructions documents can be changed and amended in time, because new methods become available or to improve their clearness. For this reason, it is crucial to keep track of the versions and differences.

CONTENTS

INTRODUCTION

The following document describes the accepted formatting and file structure and name for the submission of continuous eddy covariance raw data in ICOS. This group of variables includes high-frequency measurements of wind speed, sonic temperature and gas concentrations collected by the sonic anemometer Gill HS-50 or HS-100 and the IRGA LICOR LI-7200 or LI-7200RS. Non-CO_2 gases are not included in this section.

Summary on implementations and hardware setup for data acquisition

All the information on how to collect the measurements following the ICOS standard is reported in the relative instruction document on EC turbulent fluxes. Here a summary is made in order to contextualize the file format information, but it does not constitute an instruction document for actual implementation. Eddy covariance raw data have to be acquired digitally at a frequency of 10 Hz or 20 Hz (exceptions for few variables and specific cases possible if agreed with the ETC). Data streams from the two instruments (sonic anemometer and IRGA) have to be properly synchronized.

Although it is preferable and suggested to create a single file with both data streams, it is also possible to keep them in two separated files, one per each instrument (but same logging system). In this case the ETC will take care of properly merging them.

This Instruction document is based on the following ICOS Ecosystem protocols and documents:

- Eddy Covariance, IRGA, Tower setup and Sonic protocols and related instructions;

- Discussions between MSA and ETC;

- Requirements on data collection, traceability, processing, standardization and distribution in the Data Life Cycle, Data Policy and Data Licence.

FILE FORMAT

High frequency EC raw data

Two formats are accepted for EC raw data files: ASCII and binary. Each file must contain 30 minutes of high frequency measurements. The order of the variables is not relevant but a header must be submitted reporting the variable names according to the list reported in Tables 1 (note: for the updated version of the tables always refer to the version online in the ICOS ETC portal: www.icos-etc.eu). All the variables reported in Table 1a ("Mandatory") must be submitted otherwise the file will not be imported. Table 1b describes "Optional" variables that can be submitted and in this case they will be stored in ICOS repository. Variables with codes different from the ones reported in Table 1a and 1b will be stored in the repository at the Carbon Portal, but they will be recognised as non-ICOS variables, and will be ignored during QAQC and processing. Units of the variables must be those indicated in Table 1a and 1b.

Each row in the file must contain a single record, that means a total of 18000 expected rows in case of 10Hz data and 36000 expected rows in case of 20Hz data. Please note that this means that missing records are possible, but must be made of rows of missing values (see below), except for the timestamp. For binary files the concept of row should be interpreted as record in the matrix. Timestamp indication for each row is mandatory, unless all the used high-frequency instruments are recorded in the same file (sonic and gas analysers variables). Timestamp must report Local Standard Time (without Daylight Saving Time). The formatting of timestamp and missing data is different between ASCII and binary files, and is described in the respective sections (see below). For ASCII format, the minimum number of digits must correspond to the resolution of the sensor, as found in the instruments manual. The data format must be consistent in time: all the characteristics selected by the PI among those listed in the present document cannot be changed, unless agreed with the ETC.

Special case when external pump is used

In case an external pump is used instead of the LI-COR Flow Module, the variables related to the pump (FLOW_VOLRATE, FLOW_PRESS, FLOW_STDRATE and FLOW_PUMPDRIVE, see Tables 1a and 1b) can be acquired, after agreeing this with the ETC, at a higher sampling interval (1, 5, 10, 15 or 20 seconds) and submitted in a separate file. It is however recommended when possible to log all the variables at the same high-frequency (10 or 20 Hz)

Raw file name

The raw high-frequency data filename must follow the structure:

$$CC\text{-}\#\#\#\_EC\_YYYYMMDDHHMM\_LLN\_FFN.zzz$$

where:

- *CC-###* is the official station code (case sensitive)
- *YYYYMMDDHHMM* is the time stamp indicating the end of half hour
- *LN* (2 digits) is the logger number (each logger must be identified by a unique number)
- *FN* (2 digits) is the number of the file generated by the logger (in case multiple files are generated in the same logger each file must be identified by a unique number)
- *.zzz* is the file extension. Accepted extensions: .csv for ASCII uncompressed files; .zip for ASCII compressed files; .bin for binary files; .dat is accepted for both ASCII and binary formats.

Example of raw filename

The EC high frequency data file acquired between 16:00 and 16:30 of December 25th 2015 at the BE-Bra station, using the logger number 2 that produces only this file, in ASCII format, compressed before submission, will be named:

BE-Bra_EC_201512251630_L02_F01.zip

Please note that in the case a file is incomplete for whatever reason, the timestamp must report the end of the expected half-hour: e.g. if data acquisition stops at 15:22, the file must be timestamped _YYYYMMDD1530_. In case two files are created in one half hour, the one with the expected timestamp will be considered (however not processed if incomplete).

External header file name

An external header file must be submitted directly to the ETC using the PI Area for each type of file (see description below), even if a header is present in the raw data files. This file must have the following name structure:

$$CC\text{-}\#\#\#\_ECHEADER\_YYYYMMDDHHMM\_LLN\_FFN.csv$$

where all the parameters are the same as for the raw data filename except for the timestamp, which instead must indicate the timestamp of the first complete file where the new header applies. Whenever the structure of the raw data files will change, a new header must be submitted. Header file format can only be in comma separated value format with .csv extension.

> _Example of raw filename_
>
> On December 26th at 09.16 a.m. the columns of CO2_DRY and H2O_DRY are switched in the raw data files at BE-Bra station (see example above). A new header file must be sent before midnight through the ETC file submission system, whose name will be:
>
> BE-Bra_ECHEADER_201512261000_L02_F01.csv
>
> With the timestamp indicating the timestamp of the first complete file produced with the new structure.

When sending a new header file, there is no need to map again all the variables in the data file via the BADM, but only those that have been added, if any. Any other information that changed must be promptly communicated via the BADM system (e.g. removal of sensors, change of position/height, etc.).

ASCII format specifications

ASCII files must be comma-separated, possibly individually compressed (each half hour a file, *.zip) to save transfer time (but not mandatory). They may or may not contain one or more header lines, one of them reporting the variable names in the order they are reported in the raw files. For each type of file, i.e. for each combination of logger and file numbers (LN and FN in the filename), a header must be sent as a separate ASCII, comma-separated file, independently from the choice of the PI to embed it or not also in the file. This external header file will be composed of one line only, listing the variables in the order they are reported in the raw files. ICOS variables, i.e. variables that will be processed in the ICOS processing scheme, must be labelled according to Tables 1a (mandatory variables) and 1b (optional variables). Variables labelled differently will be stored in the repository, but they will be ignored in the processing. ICOS variables must be identical (same position and label) in the external and internal headers, when the latter is present. The external header file must only be sent before (in the same day of) the first data submission event, and then only when a change will occur, in the same day when it becomes valid (before midnight). The (new) header will be considered valid for all the data files received starting from the date and time reported in the header filename, and until replaced with a new one. Fields in both the external and, if present, in the internal header must be put between double quotes (" - ASCII code: 34). Both the header and the data must be comma-separated (, - ASCII code: 44), with the end of line indicated by a combination of carriage return (CR - ASCII code: 13) and line feed (LF - ASCII code: 10), and the dot (. - ASCII code 46) as decimal mark. Information on the raw file structure must be reported to the ETC via the BADM system, including file number, logger number, number of header-lines (if embedded) and header line number reporting the variable names (if embedded). The timestamp must be in the first column (variable name TIMESTAMP) for each record in the numeric format YYYYMMDDHHMMSS.SS with the decimal part starting with 00.10 (10Hz) or 00.05 (20 Hz) and ending always with 00.00. Alternatively, a string between double quotes following the same structure will also be accepted. Missing data will be accepted either as -9999 (without decimal digits, i.e. -9999.000 not accepted), NaN (numeric) or "NaN" (string between double quotes).

> *Example of time stamp*
>
> A 10 Hz sensor acquiring data on November 24th 2015, starting at 16:00 and ending at 16:30, will have:
>
> First time stamp: 20151124160000.10 or "20151124160000.10"
>
> Last time stamp: 20151124163000.00 or "20151124163000.00"

Binary format specifications

The binary files may or may not have a comma-separated ASCII header included, made of two or more lines with one of them reporting the variable names in the order they are reported in the raw files and one the corresponding data type. Data will be reported below as sequences of binary values. Binary data can be only of the types: unsigned longword for the timestamp, and 4-byte standard floating point (standard IEEE-754) for the data. This latter format consists of a single sign bit, an eight-bit binary exponent and a 23-bit significant. For each type of file, i.e. for each combination of logger and file numbers (LN and FN in the filename), a header must be submitted in a separate ASCII, comma-separated file, independently from the choice of the PI to embed it or not also in the file. This external header file will be composed of 2 lines, the first listing the variables in the order they are reported in the raw files, the second indicating the corresponding data type. ICOS variables, i.e. variables that will be processed in the ICOS processing scheme, must be labelled according to Tables 1a (mandatory variables) and 1b (optional variables). Variables labelled differently will be stored in the repository, but they will be ignored in the processing. ICOS variables must be identical (same position and label) in the external and internal headers, when the latter is present. Labels for the data type must be "U_LONG" for the unsigned longword and "IEEE_4" for the 4-byte standard floating point: different data type are not allowed. The header file must only be sent before (in the same day of) the first data submission event, and then only when a change will occur, in the same day when it becomes valid (before midnight). The (new) header will be considered valid for all the data files received starting from the date and time reported in the header filename, and until replaced with a new one. Fields in the header must be between double quotes (" - ASCII code: 34) and comma separated (, - ASCII code: 44). The end of line must be the combination of carriage return (CR - ASCII code: 13) and line feed (LF - ASCII code: 10). Information on the raw file structure must be reported to the ETC via the BADM system, including file number, logger number, number of header-lines (if embedded), header line number reporting the variable names (if embedded) and header line number containing the data type (if embedded). The timestamp must be split in two columns, the first reporting the seconds from an epoch date, the other the nanoseconds (variable names: TIMESTAMP_DLSS and TIMESTAMP_DLNS respectively) for each record. The epoch date and time zone must be reported in the BADM. Missing data must be as defined in the IEEE-754 standard: the infinity is represented with an exponent made of all 1 bits, and the fraction made of all 0 bits. NaN is also represented by an all-1-bits exponent, but the fraction will be composed of anything except all 0 bits. For both, the sign will be 0 for positive and 1 for negative sign.

Note: TOB1 files from Campbell Scientific dataloggers are compliant with the scheme above, provided that they are set correctly (e.g. with the reported data types)

Sonic anemometer heating

In the special and optional case of heating being applied to the sonic anemometer, the submission of the information on the time when the heat is switched on is mandatory. Two formats are possible to submit this information.

The first system consists in a daily ASCII file (.csv) reporting a flag to indicate whether the heat is on (1) or off (0). The temporal resolution of this file must be 1 minute. The file must be composed by two columns: the first column must report the timestamp in the format described in the previous sections (YYYYMMDDHHMM), and the second column the flag. The separator in the file will be the comma (, - ASCII code: 44), with the end of line indicated by a combination of carriage return (CR - ASCII code: 13) and line feed (LF - ASCII code: 10). The expected length of the file will be 1440 rows. The file must contain no header. Missing data will be accepted either as -9999 (without decimal digits, i.e. -9999.000 not accepted), NaN (numeric), or "NaN" (string between double quotes).

The second option is a daily ASCII file (.txt) reporting the info on the time the heat was switched on and the time it was switched off. It must be composed of two columns: the first column must report the variable name and the second the date and time information. In particular the variable can be either "START" or "END" while date and time must be reported using the usual format (resolution 1 minute: YYYYMMDDHHMM). The separator in the file will be the comma (, - ASCII code: 44), with the end of line indicated by a combination of carriage return (CR - ASCII code: 13) and line feed (LF - ASCII code: 10). The file must contain no header.

Example of sonic heat file – second option

On January 28th 2017 the sonic heating was switched on at 02:00 am until 07:30 am, then again from 04:43 pm all the night long until 0800 am of the 29th. The file of the 28th will be:

START,201701280200
END,201701280730
START,201701281643

and the file of the 29th will begin with a DATE_END event:

END,201701290800

Whatever the format, this file has to be submitted together with the high-frequency raw data files whenever the heating of the sonic was used. The filename will be:

$$CC\text{-}###\_SAHEAT\_YYYYMMDD\_LLN\_FFN.zzz$$

where all the parameters are the same as for the raw data filename, except the timestamp, which must indicate the day this data has been collected, and the extension, which can be csv for the first option and txt for the second.

Metadata and non-continuous information

Metadata and variables relevant for the high frequency EC measurements must be reported using the BADM system. These include the description of all the sensors used, their characteristics, setup and position, the mapping between sensor and variables produced including the logger identification and all the calibration, maintenance and disturbance events. More information about the BADM system, including how to fill and use it and the list of the metadata to be submitted are reported in the relative instruction documents

Table 1a: MANDATORY variables codes, units and descriptions. For the LI7200 two names are reported following the sensor manual (second printing, Oct. 2016): the one to the left refers to the labels as reported in the GUI to select the variables to log; the one to the right reports the labels as in the (Outputs…) Command of the Configuration Grammar, used by the sensor to communicate via Ethernet and RS232

	ICOS label	Name in the sensor manual		Units	Description
	TIMESTAMP	-		date	Time in the form YYYYMMDDHHMMSS.ss (only ASCII files)
	TIMESTAMP_DLSS	-		seconds	Seconds since the epoch reference date (only Binary files)
	TIMESTAMP_DLNS	-		nanoseconds	Nanoseconds from the beginning of the second (only Binary files)
Gill HS-50/100	U	Wc1 (UVW format)		m s^{-1}	First horizontal wind component (x)
	V	Wc2 (UVW format)		m s^{-1}	Second horizontal wind component (y)
	W	Wc3		m s^{-1}	Vertical wind component (z)
	T_SONIC	C		K	Sonic temperature
	SA_DIAG_TYPE	StaA		adimensional	Status type indicator (error, inclinometer, configuration)
	SA_DIAG_VALUE	StaD		hexadecimal value	Status value and information
LI-7200	CO2_CONC	CO2 (mmol/m3)	CO2D	mmolCO_2 m^{-3}	CO2 concentration density
	H2O_CONC	H2O (mmol/m3)	H2OD	mmolH_2O m^{-3}	H2O concentration density
	CO2_POW_SAM	CO2 Sample	CO2AW	adimensional	CO2 raw power in the sample wavelength
	H2O_POW_SAM	H2O Sample	H2OAW	adimensional	H2O raw power in the sample wavelength
	CO2_POW_REF	CO2 Reference	CO2AWO	adimensional	CO2 raw power in the reference wavelength
	H2O_POW_REF	H2O Reference	H2OAWO	adimensional	H2O raw power in the reference wavelength
	CO2	CO2 (umol/mol)	CO2MF	umolCO_2 mol^{-1}	CO2 molar fraction (in humid air)
	H2O	H2O (mmol/mol)	H2OMF	mmolH_2O mol^{-1}	H2O molar fraction (in humid air)
	CO2_DRY	CO2 dry (µmol/mol)	CO2MFd	umolCO_2 mol^{-1}	CO2 dry mole fraction (in dry air)
	H2O_DRY	H2O dry (mmol/mol)	H2OMFd	mmolH_2O mol^{-1}	H2O dry mole fraction (in dry air)
	T_CELL	Cell Temperature (°C)	AvgTemp	degrees C	Temperature of the measurement cell
	T_CELL_IN	Temperature In (°C)	TempIn	degrees C	Cell temperature where the samples enters the cell
	T_CELL_OUT	Temperature Out (°C)	TempOut	degrees C	Cell temperature where the samples exits the cell
	PRESS_CELL	Total Pressure (kPa)	Pres	kPa	Pressure in the measurement cell
	PRESS_DELTA	Head Pressure (kPa)	Dpres	kPa	Difference of pressure between cell and box
	FLOW_PRESS	Flow Pressure (kPa)	FlowPressure	kPa	Flow pressure in the sampling line
	FLOW_VOLRATE	Flow Rate (lpm)	VolFlowRate	liters min^{-1}	Volume flow rate in the sampling line
	GA_DIAG_CODE	Diagnostic value	DiagVal	adimentional	diagnostic value (will be converted by ETC)
	GA_TUBE_HEAT	*(to be added in the sensor manual)*		W	Watts provided to the tube for heating
	GA_TUBE_HEAT_V	*(to be added in the sensor manual)*		V	Volts provided to the tube for heating

Table 1b: OPTIONAL variables codes, units and descriptions. For the LI7200 two names are reported following the sensor manual (second printing, Oct. 2016): the one to the left refers to the labels as reported in the GUI to select the variables to log; the one to the right reports the labels as in the (Outputs…) Command of the Configuration Grammar, used by the sensor to communicate via Ethernet and RS232

	ICOS label	Name in the sensor manual		Units	Description
Gill HS-50/100	WS	Wc2 (Polar format)		m s^{-1}	Wind speed
	WD	Wc1 (Polar format + POLARWRAP 360)		degrees	Wind direction (0 to 360 degrees)
	WD_H	Wc1 (Polar format + POLARWRAP 540)		degrees	Wind direction with hysteresis (0 to 540 degrees)
	W_AX1	Wc1 (Axis Speeds)		m s^{-1}	Wind component in the direction of sonic axis 1
	W_AX2	Wc2 (Axis Speeds)		m s^{-1}	Wind component in the direction of sonic axis 2
	W_AX3	Wc3 (Axis Speeds)		m s^{-1}	Wind component in the direction of sonic axis 3
	SA_CHKSUM	Cc		adimensional	Checksum diagnostic value on SA data transmission
	T_ABS	T		K	Absolute temperature from Platinum Resistance Thermometer (PRT)
LI-7200	CO2_MASS	CO2 (mg/m3)	CO2MG	mgCO$_2$ m^{-3}	CO2 mass density
	H2O_MASS	H2O (g/m3)	H2OG	gH$_2$O m^{-3}	H2O mass density
	CO2_ABS	CO2 Absorptance	CO2Raw	adimensional	CO2 absorptance
	H2O_ABS	H2O Absorptance	H2ORaw	adimensional	H2O absorptance
	FLOW_STDRATE	Flow Rate (slpm)	MeasFlowRate	liters min^{-1}	Flow rate in the sampling line standardised for T and PRESS
	T_DEW	Dew Point (°C)	DewPt	degrees C	H2O as dew point temperature
	T_GA	Block Temperature (°C)	Temp	degrees C	Temperature of gas analyser
	PRESS_BOX	Box Pressure (kPa)	Apres	kPa	Pressure of the box containing the logger and the electronics of the GA
	FLOW_PUMPDRIVE	Flow Drive (%)	FlowDrive	%	Drive input to the pump
	CO2_STR	CO2 Signal Strength	CO2SS	adimensional	Signal strength of CO2
	H2O_STR	H2O Signal Strength	H2OSS	adimensional	Signal strength of H2O
	DELTA_STR	Delta Signal Strength	DeltaSS	adimensional	Difference between signal strength of two gases
	GA_CHKSUM	checksum		adimensional	check integrity of data