

ISFS NETCDF FILE CONVENTIONS

Steve Oncley (ISFS Lead Scientist) Jacquelyn Witte (ISF Data Manager)

Integrated Surface Flux System (ISFS)
Earth Observing Laboratory / In-situ Sensing Facility
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH
PO Box 3000, Boulder, CO 80307-3000 USA

Document Version Control

Version	Date	Author	Description
2.0	June 2020	Witte	Assembly and content editing
1.0		Oncley	Author, original web content



Table of Contents

Motivation	3
EOL/ISF Data Management Tenet	3
Considerations	3
Software	3
File name convention	3
Time Sampling and Dimension	4
Station Dimension	4
Time Representation	5
Variable Names and Attributes	5
Data Sampling Time-Averaged Data Resampled High Rate Data	7 7 7
Coordinate Systems of Wind Variables	8
Quality Indicators of 3D Wind Variables from CSAT3 Anemometers	9
Counts Variables	10
Metadata - ISFS Global Attributes	11
List of ISFS Variables	11
Meteorological Variables	12
Sonic Anemometer Variables	13
CO2 and H2O Open-Path Gas Analyzer Variables	14
Radiation Variables	15
Soil variables	16
Precipitation Variables	17
Engineering Variables	18
Names of Covariances and Higher Moments:	19



Motivation

This is a readme guide to understanding how ISFS netCDF data files are constructed and how its variables are defined. This is an amalgamation of a number of ISFS webpages that have been written by Steve Oncley. The objective is to consolidate all this information into a document that (i) has been identified as a long-lasting location (Google Shared Drive: ISF Data Management), (i) avoids dependency on web pages that are subject to altered links and deletion, and (iii) becomes part of the group of ISFS data QC documentation that accompanies datasets uploaded to the EOL Archive project pages.

EOL/ISF Data Management Tenet

Management and oversight of ISF data assets requires developing an enduring, descriptive, and practical library of foundational documents that preserve critical metadata for past and future analysis. Guides such as these inform data users on interpretation and better understanding of EOL/ISF data holdings.

Considerations

ISFS follows the CDL (Common Data Language) notation for netCDF and CF conventions, where defined. This is a living document subject to updates and changes as EOL keeps up with data format developments.

1. Software

ISFS data provided in netCDF format. Information on the NetCDF file format and software is available at http://www.unidata.ucar.edu/software/netcdf/. Support for reading NetCDF files is available for software written in C, C++, Fortran, Python, Java, MATLAB, R and IDL.

An 'isfs' R package, specifically written for reading and analyzing ISFS data in R, is available at https://github.com/NCAR/R-eol/tree/master/packages/isfs

2. File name convention

Files are grouped by a fixed time length, typically 24 hours, but sometimes longer or shorter depending on the time resolution of the data.

The file names for 24 hour files will typically be of the form *isfs_prefix_YYYYMMDD.nc*. *YYYY* is the year, and *MMDD* is the 2 digit month and day. *prefix* is a unique, short descriptor of the dataset that distinguishes it from other ISFS datasets with the same time stamp. The files will contain data from 00:00 to 24:00 UTC for the given day.

High-resolution data files have time series that are fractions of a day, i.e. hourly, 2-hr, 4-hr, etc. File names will be in the form isfs_prefix_hr_YYYYMMDD_HH.nc, where prefix is a short



descriptor of the dataset, and HH is the two-digit hour in UTC of the start time of the data in the file.

3. Time Sampling and Dimension

ISFS variables will have a corresponding time dimension. For example, a variable that is stored at 1 sample/s in a 24 hour netCDF file will have a time dimension of 24hr x 60s x 60min = 86400 samples, for the number of seconds in a day. 24 hour files containing 5 minute averaged variables will have a time dimension of 86400/300=288 samples/5 min.

The standard ISFS dataset contains 5 minute averaged variables and statistics. However, depending on the project needs, datasets may be provided with differing time resolution. High resolution datasets will contain variables at higher time resolution typically 1/20th of a second - meaning 20 samples/s variables will have a sample dimension of 20 in addition to the time dimension:

```
netcdf isfs_geo_hr_20190920_15 {
dimensions:
    time = UNLIMITED ; // (3600 currently)
    sample = 20 ;
```

If there are more than one sample dimension in a file, the additional sample dimensions will have unique names, with the sampling rate as a suffix, for example, sample_10 for 10 sample/second variables.

4. Station Dimension

The station dimension applies to variables *altitude*, *latitude* and *longitude*. These variable names have been sampled at more than one ISFS location and have a station dimension whose value will be the number of ISFS stations that were deployed at the project:

```
netcdf isfs_tc_20190210 {
dimensions:
    station = 15 ;
    stationnamelen = 4 ;
    time = UNLIMITED ; // (288 currently)
variables:
    char station(station, stationnamelen) ;
    double latitude(station) ;
        latitude:units = "degree_N" ;
        latitude:short_name = "latitude" ;
        latitude:_FillValue = 1.e+37 ;
        latitude:long name = "Degrees N Latitude" ;
```



5. Time Representation

The base_time variable contains one value, the time of the start of the file, as a number of POSIX (non-leap) seconds since 1970 Jan 1, 00:00 UTC. Refer to http://en.wikipedia.org/wiki/Unix time for more information about POSIX or Unix time.

The files also contain a *time* variable with a time dimension whose values represent the number of seconds since the start of the project. The *time* variable has a *unit*s attribute which provides a human-readable representation of *base_time*, as seen in the output below from a Relampago project file:

6. Variable Names and Attributes

The primary components of an ISFS netCDF file are variables and its attributes. Each variable stored will have an attribute. ISFS uses the CDL (Common Data Language) notation for defining an attribute as

```
variable_name:attribute_name = list_of_values ;
```

where attribute_name will contain one or more of the following

- _FillValue = Typically 1.e+37f, indicating that a measurement is not available at the given time.
- long name = A more complete description of the measurement or sensor.
- short_name = The name of the variable, as assigned by the ISFS data acquisition system.
- counts = If a variable is a time-averaged mean or higher moment, the counts attribute will indicate the name of the netCDF variable containing the number of data points that were averaged.
- *units* = A string indicating the measurement units.

The ISFS variable names, as indicated in the *short_name* attribute, often contain periods "." to separate fields that are typically defined in the order of *name*, *height of the measurement sensor*, and *site name* (defined in the ISFS QC report). Single quote marks (the 'symbol) indicate a derived variable, i.e. moments.



Few example below taken from the CHEESEHEAD project:

Software to read netCDF files access the variable name denoted by underscores. Note, netCDF variable names cannot contain periods or quote marks. The exception is when using the 'isfs' R package whose reader program accesses data via the *short_name* attribute (denoted by periods). This is because the ISFS data acquisition system defines its variable names with periods. When producing netCDF files, the periods are substituted for underscores and the original name is kept as a *short_name* attribute. Thus, in a netCDF file the soil temperature at 1.9 cm depth located at site nw1 is defined by ISFS as Tsoil.1.9cm.nw1 and by netCDF as Tsoil_1_9cm_nw1:

A covariance between u_30m_nw1 and h2o_30m_nw1 - see examples above - would have an ISFS variable name and short_name attribute of u'h2o'.30m.nw1, and a NetCDF variable name of u_h2o__30m_nw1:

The ISFS variable names are listed at the end of this guide.



7. Data Sampling

The ISFS DAS acquires samples from sensors in an asynchronous manner, with each sensor having its own internal processor clock which is not synchronized with any other sensors or any external clock source. The samples from each sensor are time-tagged at the time of their receipt by the data system, based on an accurate system clock. This system clock is continually monitored and adjusted by NTP software, using a GPS reference clock with a precise PPS (pulse-per-second) signal, and is generally correct to within 50 microseconds of the PPS signal.

During post-processing, time tags for samples from some sensors are further adjusted as appropriate for the given sensor, based on documentation of the internal delays of the sensor.

NetCDF files are merged in time by sampling onto an even time grid, either by time-based averaging, or by a simple nearest-in-time resampling algorithm. 5 minute averages, for example, use a time-based averaging approach. High rate sampling uses the nearest-in-time approach.

7.1. Time-Averaged Data

For averaged data, each time corresponds to the middle of each averaging period. Typically for time-averaged data, there will be only one time resolution in the file and no variables will have a sample dimension. For example, for 5 minute (300 second) averages, the values of the time variable will be 150, 450, 750, etc, which are the middle times of each successive 300 second interval, as a number of seconds since the base-time.

7.2. Resampled High Rate Data

Before being written to the NetCDF files, the raw, asynchronous samples are re-sampled to an evenly spaced time sequence, using a simple method of matching the raw sample nearest-in-time to the evenly-spaced times. No interpolation or averaging is done.

The time-tag of a sample with time index i, for a variable without a sample dimension, is simply:

 $t_i = base time + time_i$

This time-tag represents the time of the sample as a number of POSIX seconds since Jan 1, 1970 00:00 UTC.

For variables with a sample dimension, the samples corresponding to a given time index are evenly spaced around the corresponding time_i. The time-tag for a sample from a variable that was re-sampled at rate R/s, with time index i, sample index i, is:



```
t_{i,j} = base\_time + time_i - (dT / 2) + (dT / sample) \times (1/2 + j) dT = sample / R \ seconds, \ is \ the \ interval \ between \ time_i \ values i \ is \ the \ time \ index j \ is \ the \ sample \ index, \ ranging \ from \ 0 \ to \ (sample - 1)
```

The sample dimension for a variable is usually chosen to be its rate, R, so that dT = 1 second.

8. Coordinate Systems of Wind Variables

Wind measurements from 2D anemometers are typically provided in geographic coordinates, where the direction the wind is coming from is reported with respect to 0=north, 90=east, etc. When wind is reported in terms of its U and V components, +U is wind flowing to the east, and +V is wind to the north. Wind measurements from 3D sonic anemometers are also typically in geographic coordinates for time averaged and high rate netCDF files. Correction of sonic tilt will depend on the type of correction and this information can be found in the project QC report. High rate datasets are not tilt corrected.

Current ISFS netCDF files contain some information on the wind coordinates in the global attributes section. The ncdump program can be used to display the global attributes. For example, a file containing 3D wind data that have been rotated from instrument to geographic coordinates, but have not been tilt-corrected, has these global attributes:

```
// global attributes:
    :dataset = "qc_geo_notiltcor";
    :dataset_description = "QC, winds in geographic, non-tilt
corrected coordinates";
    :wind3d_horiz_coordinates = "geographic";
    :wind3d_horiz_rotation = 1;
    :wind3d_tilt_correction = 0;
```

The dataset, dataset_description, and wind3d_horiz_coordinates attributes are descriptive.

wind3d_tilt_correction is a logical value, 0=false, 1=true, indicating whether tilt correction has been applied to the 3D winds.

wind3d_horiz_rotation is a logical value, which if 0 (false), indicates the 3D wind data have not been rotated, and hence are in instrument coordinates.

A value of 1 (true) for wind3d_horiz_rotation indicates the winds have been rotated in the horizontal plane, and the wind3d_horiz_coordinates attribute will indicate the resultant coordinates.

In ISFS processing, a tilt correction is first applied to the sonic winds, followed by the horizontal rotation.



For older files, without the above global attributes, the 3D wind coordinates should be documented in the project page. Hints may also be found in the file names, with abbreviations such as "geo" (geographic) or "instr" (instrument), and "tc" for tilt correction or "ntc" for no tilt correction.

9. Quality Indicators of 3D Wind Variables from CSAT3 Anemometers

ISFS 3D wind measurements are commonly done with Campbell Scientific CSAT3 sonic anemometers. With every wind sample, a CSAT3 also outputs a diagnostic value, consisting of four bit fields, where a bit will be set if a problem is detected. As documented in the CSAT3 manual (https://s.campbellsci.com/documents/us/manuals/csat3.pdf) in table B7, "Decoding the Diagnostic Flags from Word 4", the bits have the following meaning:

CSAT3 bit	Indication	ISFS bit	ISFS decimal value
b12	Sonic signal amplitude too low	0	1
b13	Sonic signal amplitude too high	1	2
b14	Poor signal lock	2	4
b15	Difference in the speed of sound between the three non orthogonal axes is greater than 2.360 m s–1 (~ 4 °C @ 25 °C)	3	8
	Unexpected value for sample counter: possible data loss	4	16

In high rate ISFS netCDF files this diagnostic is usually stored in a variable called diagbits. For example, a sonic at 20 meters with wind variables having short names of u.20m, v.20m, w.20m and tc.20m, will also have a variable with a short name diagbits.20m.

As shown in the table, CSAT3 bits 12-15 are shifted to the least significant bits (0-3) before being stored in diagbits. The decimal value of diagbits for a sample will be the sum of the shown decimal values.

The most common cause of a non-zero diagbits is the presence of water droplets on the anemometer transducers.

The u, v, w and to values of a high rate 3D wind sample are set to the missing value by ISFS processing if any of bits 0-3 (CSAT3 bits b12-b15) are non-zero.

Bit 4 in diagbits is added by ISFS software to indicate that the sample counter provided by the CSAT3 was not the expected value (the previous counter plus 1, modulus 64). If diagbits has a value of 16 or greater, it is an indication samples are being lost during sampling, transmission or storage. The CSAT3 sample counter is not otherwise stored in the NetCDF data.



Since the high rate values are set to the missing value when bits 0-3 are non-zero, means and covariances of CSAT3 wind data will not include those flagged high rate samples. NetCDF files of averages generally include an Idiag variable for each sonic anemometer. Idiag is an average of a logical diagnostic over the period, with value of 0 or 1 indicating whether diagbits was non-zero. The average of Idiag is then the fraction of time over each average period that diagbits was non-zero.

To screen averaged data, choose a maximum allowed value for ldiag, for example 0.01, and discard the means or covariances for a sonic when ldiag exceeds that value.

10. Counts Variables

If an averaged variable has a counts attribute, it will be the name of a variable containing the number of points in each average. For example, a 5 minute average of a sonic running at 20 sample/sec, would have a counts value of 6000 (+-1) for a period of clean data. The counts variable can also be used to screen suspect averages.

11. Metadata - ISFS Global Attributes

Metadata are stored as global attributes at the end of each ISFS netCDF file. Documentation, particularly in the recent projects, will contain the following information below.

history : File creation date and time and revision control information

NIDAS_version : Version of NIDAS configuration

calibration_file_path : Local directory of the calibration files for ISFS bookkeeping dataset : Abbreviated descriptor of the dataset, e.g qc_geo_tiltcor

dataset_description : Full description of the dataset, e.g. . QC, tile corrected,

geo-referenced

project config :Local directory of the project configuration for ISFS bookkeeping

wind3d_horiz_coordinates : Reference sonic anemometer winds in either geographic or

instrument coordinates

file length seconds :Number of samples

wind3d_horiz_rotation :Adjustment to the sonic anemometer winds in the direction of

geometric north =1(yes) = 0 (no)

wind3d tilt correction :Sonic anemometer winds adjusted to align parallel to the ground

=1(yes) = 0 (no) = 1

12. List of ISFS Variables



Many meteorological variables are sampled with variable time responses. Names of variables whose sample rates are once per second or less begin with an upper case character, for example P, T, RH, Dir, Spd, U, and V. Variables sampled at higher rate (usually 10/sec or more) are in lower case. Lower case variables are used for eddy correlation calculations, for example: t, u, v, w, and q. Derived fluxes use the abbreviation commonly used in the literature, which may be upper or lower case, for example, H for sensible heat flux.

ISFS netCDF files may contain, but are not limited to the following list of parameters, divided by category. Note that a few of the variables are present for older datasets (PAM projects era), as noted in the the source/notes column.

ISFS applies a fill value of 1.e+37f to all netCDF variables.

12.1. Meteorological Variables

Variable name	unit	long name	<instrument> source or notes</instrument>
Р	mb	Barometric Pressure, <instrument></instrument>	Vaisala PTB 210, Paroscientific 6000
Т	degC	Air Temperature, <instrument></instrument>	NCAR hygrothermometer
RH	%	Relative humidity	NCAR hygrothermometer
T _C	degC	Sonic temperature, Gill WindObserver	2D Gill WindObserver
Spd	m/s	Wind speed	2D Gill WindObserver
Dir	deg	Wind direction	Gill WindObserver, 0=North, 2D 90=East
U	m/s	Wind U component	+ is from the West
V	m/s	Wind V component	+ is from the South
T _V	degC	Dew point temperature	in older datasets
T _W	degC	Wet bulb temperature	in older datasets
Q	g/kg moist air	specific humidity	in older datasets
MR	g/kg dry air	Mixing ratio	in older datasets



12.2. Sonic Anemometer Variables

Variable name	unit	long name	Source or notes
u	m/s	Wind U component, CSAT3	Sonic anemometer
V	m/s	Wind V component, CSAT3	Sonic anemometer
w	m/s	Wind W component, CSAT3	Sonic anemometer
spd	m/s	Wind speed	Sonic anemometer
dir	deg	Wind direction	Sonic anemometer
Idiag	none	CSAT3 logical diagnostic, 0=OK, 1=(diagbits!=0)	Sonic anemometer
Vazimuth	deg	pointing direction (0=N,90=E) of sonic +V axis	in older datasets, hand held compass,
uflag, vflag, wflag, tcflag	none	sonic spikes	in older datasets
diag	none	Sonic logical diagnostic	in older datasets, logical value, 0=all diag bits are zero, 1=one or more non-zero diag bits

12.3. CO₂ and H₂O Open-Path Gas Analyzer Variables

Variable name	unit	long name	Source or notes
t _C	degC	Virtual air temperature from speed of sound, CSAT3	Sonic anemometer
h2o	g/m^3	Water vapor density from CSI IRGA	EC150 H2O/CO2 infrared gas analyzer
co2	mg/m^3	CO2 density from CSI IRGA	EC150 H2O/CO2 infrared gas analyzer
Pcell, Pirga	mb	CSI IRGA pressure	Li7500, EC150; will have dynamic pressure error due to poor port placement
Tirga	degC	CSI IRGA temperature	diagnostic flag for EC150 H2O/CO2 infrared gas analyzer
irgadiag	none	CSI IRGA diagnostic	diagnostic flag for EC150 H2O/CO2 infrared gas analyzer
t	degC	air temperature	in older datasets
kh2o	g/m^3	water vapor density	in older datasets, krypton hygrometer, not corrected for absorption by O ₂



q	g/kg of moist air	Specific humidity	in older datasets
mr	g/kg of dry air	Water vapor mixing ratio	in older datasets

12.4. Radiation Variables

Variable name	unit	long name	Source or notes
Rsw.in	W/m^2	Incoming Short Wave, <instrument></instrument>	Hukesflux NR01
Rsw.out	W/m^2	Outgoing Short Wave, <instrument></instrument>	Hukesflux NR01
Rpile.in	W/m^2	Incoming Thermopile, <instrument></instrument>	Hukesflux NR01
Rpile.out	W/m^2	Outgoing Thermopile, <instrument></instrument>	Hukesflux NR01
Tcase	degC	Case temperature, <instrument></instrument>	Hukesflux NR01 thermister
Wetness	V	Decagon Leaf Wetness Sensor	Can detect water source from shape of signal; flat=dry
Rlw.in	W/m^2		Derived from Rpile.in, Tcase
Rlw.out	W/m^2		Derived from Rpile.out, Tcase
Rsw.sum	W/m^2		signed sum of short wave radiation, Rsw.in - Rsw.out
Rlw.sum	W/m^2		signed sum of long wave radiation, Rlw.in - Rlw.out
Tsfc	degC		Derived surface temperature
Tdome	degC		in older datasets, radiometer dome temperature, thermister



12.5. Soil variables

Variable name	unit	long name	<instrument> source or notes</instrument>
Tsoil	degC	Soil temperature	
Qsoil	vol%	Soil moisture	
Gsoil	W/m^2	Soil Heat Flux	
Gsfc	W/m^2		Heat flux at surface. Derived from Gsoil, Tsoil, Csoil
Vheat	V	<instrument> heater voltage</instrument>	Hukseflux TP01
Vpile	uV	<instrument> thermopile after heating</instrument>	Hukseflux TP01
Tau63	sec	<instrument> time to decay to 37% of Vpile.on-Vpile.off</instrument>	Hukseflux TP01
Landasoil	W/m/ DegK	<instrument> derived thermal conductivity</instrument>	Hukseflux TP01
asoil	m^2/s		Soil thermal diffusivity. Derived from TP01 Vheat, Vpile, Tau63
Csoil	J/(m^3 degC)		Soil heat capacity. Derived from TP01 asoil, Lambdasoil
Tice	degC		Ice temperature. In older datasets
Gice	W/m^2		Heat flux at a certain depth. In older datasets

12.6. Precipitation Variables

Variable name	unit	long name	<instrument> source or notes</instrument>
Rainr	mm/h	Rain rate, <instrument></instrument>	Tipping bucket, OTT Parsivel2
Tcell	degC	Sensor temperature, OTT Parsivel2	
WX	none	Present weather code, OTT Parsivel2	
Vis	m	MOR visibility, OTT Parsivel2	
N	none	Number of particles, OTT Parsivel2/	
Raina_tip	mm	Tipping bicket	Rain accumulation



12.7. Engineering Variables

Variable name	unit	long name	<instrument> source or notes</instrument>
GPSnsat	none	Number of GPS satellites tracked	
GPSstat	none	GPS rcvr status: 1=OK(A), 0=warning(V)	
GPSdiff	s		
Stratum	none	NTP stratum	Status of time service (1 is best)
Timeoffset	usec	Clock offset, system-reference	Error in data system clock
Ifan	mA	Aspiration fan current, NCAR hygrothermometer	Current consumption of TRH sensor's fan
Vbatt	V	Battery voltage	
lload	mA	Load current	
Icharge	mA	Charging current	
Tbatt	degC	Batt. Temperature	
		Relevant to older data	sets
Tbox	degC	Electronics temperature	
Vin, Vdsm	V	power supply input voltage	
lin	mA	power supply current	
V3.3	V	Voltage to 3.3V supply	
13.3	mA	current consumption of 3.3V supply	
lxb	mA	current to Xbee radio	
19x	mA	current to 9Xtend radio	
Isnsr	mA	current to sensor	



12.8. Names of Covariances and Higher Moments:

ISFS derives numerous time averaged flux statistics and it is common for a netCDF file to contain a lengthy list of eddy correlation flux parameters. Covariances calculations encompass:

- 2nd order moments, i.e. v'tc' [m/s degC], u'u' [(m/s)^2]
- 3rd order moments, i.e. v'v'v' [(m/s)^3]
- 4th higher moments, i.e. w'w'w' [(m/s)^4].

The list for ISFS derived parameters is too long to tabularize in this document and varies from project to project depending on the science team research needs. However, it is standard practice to include all possible combinations of 2nd order moments from the 3D sonic anemometers and h2o/co2 gas analyser measurements.