

DataFormat

The data files are text files stored in UTF-8. The line feed code is LF. They contain header records and data records.

1. Header record

Header records are created automatically from metadata registered by data contributors. Lengths are variable and defined at the top of the record.

Composition items are as per the table on the following pages.

➤ Data averaging

Certain types of daily and monthly data are calculated by WDCGG at the request of Contributors. The procedures for such calculation are outlined below.

- 1) Daily mean data are calculated using hourly data.
Monthly mean data are calculated using daily data.
- 2) Averages are calculated from data whose QC flag status is *1: valid (background)* or *2: valid*.
- 3) At least two data points must be available for average calculations.
If only one data point is available, the average is set to "-999.999."
- 4) Average data are based on the simple arithmetic mean.
The Value_unc of average data is the standard deviation of available data, and the Nvalue is the number of available data points.

Item name	Example	Number of digits	Description of the item
# header_lines :	187	Variable	header length
# Data Set Name:	co2_mlo_surface-flask_2_3001-9999.event	Variable	Character string that uniquely identifies the data set
# Data_Set_Version:	0002-5002-1001-01-02-3001.2018-03-09-1519	41	Version of the data set
# Data_Set_Fair_Use :	For Scientific purposes, access to these data is unlimited and provided without charge. By their use you accept that an offer of co-authorship will be made through personal contact with the contributors whenever substantial use is made of their data. In all cases, an acknowledgement must be made to the contributors and to the data centre when these data are used within a publication.		This is the WDCGG fair use statement agreed upon by data contributors.
#			
# GLOBAL ATTRIBUTES			
#			
# site_gaw_id :	MLO	3	3-letter site identification code as defined by GAW/SIS for stationary platforms
# site_gaw_type :	GAW Global	Variable	GAW station category
# site_name :	Mauna Loa (HI)	Variable	Standard site name
# site_country/territory :	United States	Variable	Country/territory in which site is located
# site_wmo_region :	REGION V (South-West Pacific)	Variable	WMO Region
# site_address1 :	P.O. Box 275 Hilo, HI 96720 HAWAII	Variable	-
# site_address2 :		Variable	-
# site_address3 :		Variable	-
# site_latitude :	19.5362300873	Variable	Latitude (decimal degree) at representative site location
# site_longitude :	-155.5761566162	Variable	Longitude (decimal degree) at representative site location
# site_elevation :	3397	Variable	Ground or surface elevation at representative site location
# site_elevation_unit :	m	1	Units are meters above sea level (masl)
# site_lst2utc :	UTC-10:00	3 or 9	Hour conversion from UTC to LST
# site_climate_zone :	Cfb	Variable	Koepfen's classification of climate in which site is located
# site_climate_zone:comment :	Warm temperate climate, fully humid, warm summer	Variable	Climatic type
# dataset_creation_date :	2018-04-05	10	Date when the data set was prepared
# dataset_parameter :	co2	Variable	Identifies trace gas species included in data set
# dataset_parameter_large_class :	Greenhouse Gas	Variable	Large classification of trace gas species included in data set
# dataset_parameter_small_class :		Variable	Small classification of trace gas species included in data set
# dataset_parameter_name_1 :	CO2	Variable	Identifies trace gas species included in data set
# dataset_parameter_name_2 :	Carbone Dioxide	Variable	Identifies trace gas species included in data set (alias name)
# dataset_parameter_name_3 :		Variable	Identifies trace gas species included in data set (alias name)
# dataset_parameter_name_4 :		Variable	Identifies trace gas species included in data set (alias name)
# dataset_parameter_name_5 :		Variable	Identifies trace gas species included in data set (alias name)
# dataset_project :	surface-flask	Variable	Typically identifies sampling platform and strategy
# dataset_platform :	fixed station	Variable	Fixed or Mobile
# dataset_selection :	All event data	Variable	Brief description of how data have been selected by data contributor
# dataset_selection_tag :	event	Variable	Short descriptor to help convey how data have been selected by data contributor
# dataset_time_zone :	UTC-10:00	Variable	This item expresses the differential to UTC.
# dataset_start_date :	1969-08-20T17:55:00Z	20 or 25	Date of first item in data set (ISO 8601 format). The letter '*' indicates that the time zone is unknown.
# dataset_end_date :	2016-12-27T20:44:00Z	20 or 25	Date of last item in data set (ISO 8601 format). The letter '*' indicates that the time zone is unknown.
# dataset_statistic_code :	1	1	Statistic creator identification number
# dataset_statistic :	contributor	Variable	Statistic creator (contributor or wdcgg)
# dataset_description :	Air samples were collected into 24 titanium flasks using manual air sampling pump.	0-3500	Brief description of data set contents by data contributor
# dataset_aim_of_observation_code :	1	1	Identification number of aim of observation
# dataset_aim_of_observation :	Background observation	Variable	Aim of observation
# dataset_buffer_code :	3001	4	Identification number necessary for additional observational condition
# dataset_buffer :	ccgg	Variable	Identification necessary for additional observational condition
# dataset_buffer:comment :	NOAA Carbon Cycle Greenhouse Gases	Variable	-
# dataset_footnote_code :	9999	4	Identification number of data set among same observational condition
# dataset_footnote :	Others	Variable	Identification name of data set among same observational condition
# dataset_footnote:comment :		Variable	-
# dataset_processing :	Uncertainty in the measurements of CO2 from discrete samples has not yet been fully evaluated. Key components of it are our ability to propagate the WMO XCO2 scale to working standards, the repeatability of the analyzers used for sample measurement, and agreement between pairs of samples collected simultaneously. Zhao and Tans (2006) determined that the internal consistency of working standards is +/- 0.02 ppm (68% confidence interval). The typical repeatability of the analyzers, based on repeated measurements of natural air from a cylinder, is +/- 0.03 ppm. Average pair agreement across the entire sampling network is +/- 0.1 ppm. The Pacific Ocean Cruise (POC, travelling between the US west coast and New Zealand or Australia) data have been merged and grouped into 5 degree latitude bins. For the South China Sea cruises (SCS) the data are grouped in 3 degree latitude bins. Historically, samples have been collected using two general methods: flushing and then pressurizing glass flasks with a pump, or opening a stopcock on an evacuated glass flask; since 28 April 2003, only the former method is used. During each sampling event, a pair of flasks is filled.	0-3500	Description of processing procedure for data set preparation
# dataset_hr_mean_processing :		0-3500	
# dataset_da_mean_processing :		0-3500	
# dataset_mo_mean_processing :	Monthly means are produced for each site by first averaging all valid measurement results in the event file with a unique sample date and time. Values are then extracted at weekly intervals from a smooth curve (Thoning et al., 1989) fitted to the averaged data and these weekly values are averaged for each month to give the monthly means recorded in the files. Flagged data are excluded from the curve fitting process. Some sites are excluded from the monthly mean directory because sparse data or a short record does not allow a reasonable curve fit. Also, if there are 3 or more consecutive months without data, monthly means are not calculated for these months.	0-3500	details of procedures to make hourly, daily and monthly mean data
# dataset_transition_flag :	0	1	Transition flag indicating dataset was made using the former WDCGG data file or not
# dataset_transition :	Dataset was made from entire historical data provided from contributor from 2018 new WDCGG website open onward		
# dataset_reference_total_listed :	3	Variable	Number indicating how many references to published literature to expect in this file
# dataset_reference_#_name :	Zhao, C., and P.P. Tans (2006). Estimating uncertainty of the WMO Mole Fraction Scale for carbon dioxide in air, J. Geophys. Res. 111, D08S09, doi: 10.1029/2005JD006003.	0-1000	Reference provided by data contributor. The letter '#' represents a number from 1 to maximum (dataset_reference_total_listed).
	Thoning, K.W., T.J. Conway, N. Zhang, and D. Kitzis, 1995. Analysis system for measurement of CO2 mixing ratios in flask air samples, J. Atmos. and Oceanic Tech., 12, 1349-1356.	0-1000	
	Thoning, K.W., P.P. Tans, and W.D. Komhyr, 1989. Atmospheric carbon dioxide at Mauna Loa Observatory 2. Analysis of the NOAA GMCC data, 1974-1985. J. Geophys. Res., 94, 8549-8565.	0-1000	
# contributor_organization_code :	2	Variable	Contributor identification number
# contributor_acronym :	NOAA	1-16	Contributor abbreviation or acronym
# contributor_name :	Earth System Research Laboratory, NOAA	1-255	Contributor name
# contributor_address1 :	R/GMD1	0-255	-
# contributor_address2 :	NOAA/ESRL	0-255	-
# contributor_address3 :	325 Broadway Boulder, CO 80305-3337	0-255	-
# contributor_country/territory :	United States	Variable	Country/territory name
# contributor_url :	http://www.cmdl.noaa.gov/index.html	0-255	URL of contributor
# contact_total_listed :	1	Variable	Number of contact persons associated with the data set
# contact_#_name :	Firstname LASTNAME	1-100	Contact person name. The letter '#' represents a number from 1 to maximum (contact_total_listed).
# contact_#_address1 :	1-3-4 Otemachi	0-255	-
# contact_#_address2 :	Chiyoda-ku	0-255	-
# contact_#_address3 :	Tokyo 100-8122	0-255	-
# contact_#_country/territory :	Japan	Variable	Country/territory name
# contact_#_organization :	Japan Meteorological Agency	1-255	Organization name
# contact_#_organization_acronym :	JMA	1-16	Acronym of organization name
# contact_#_organization_url :	http://www.jma.go.jp/jma/indexe.html	0-255	URL of organization
# contact_#_email :	name@domain	1-100	Email address of contact person
# contact_#_tel :	+81-1-2345-6789	0-50	Telephone number of contact person
# contact_#_fax :	+81-1-2345-6789	0-50	Fax number of contact person
# collaborator_total_listed :	1	Variable	Number of contributing collaborators associated with the data set
# collaborator_#_organization_code :	2	Variable	Collaborator identification number. The letter '#' represents a number from 1 to maximum (collaborator_total_listed).
# collaborator_#_acronym :	NOAA	1-16	Collaborator abbreviation or acronym
# collaborator_#_name :	Earth System Research Laboratory, NOAA	1-255	Collaborator name
# collaborator_#_address1 :	R/GMD1	0-255	-
# collaborator_#_address2 :	NOAA/ESRL	0-255	-
# collaborator_#_address3 :	325 Broadway Boulder, CO 80305-3337	0-255	-
# collaborator_#_country/territory :	United States	Variable	Country/territory name
# collaborator_#_url :	http://www.cmdl.noaa.gov/index.html	0-255	URL of collaborator
# collaborator_#_person_total_listed :	1	Variable	Number of person in collaborator organization associated with the data set
# collaborator_#_person_#_name :	Firstname LASTNAME	1-100	Collaborator person name. The letter '#' represents a number from 1 to maximum (collaborator_#_person_total_listed).
# collaborator_#_person_#_address1 :	1-3-4 Otemachi	0-255	-
# collaborator_#_person_#_address2 :	Chiyoda-ku	0-255	-
# collaborator_#_person_#_address3 :	Tokyo 100-8122	0-255	-
# collaborator_#_person_#_country/territory :	Japan	Variable	Country/territory name
# collaborator_#_person_#_email :	name@domain	1-100	Email address of collaborator person
# collaborator_#_person_#_tel :	+81-1-2345-6789	0-50	Telephone number of collaborator person
# collaborator_#_person_#_fax :	+81-1-2345-6789	0-50	Fax number of collaborator person
# ORG_QCflag_description :	Hourly data 0-6: no background condition 7: background condition 9: missing Daily and Monthly data 2: no valid hourly data 3: valid 9: missing	Variable	Original data flag set by contributor
# intake_height_total_listed :	1	Variable	Number indicating how many kinds of intake heights otherwise "See data part"
# intake_height_# :	20	1-6	Height above ground at which air sample was collected. The letter '#' represents a number from 1 to maximum (intake_height_total_listed).
# intake_height_#_units :	m	1	Units are meters above ground level (magl).
# intake_height_#_start_date :	1993-01-01T00:00:00+09:00	20 or 25	Date of first item in data during the intake height period (ISO 8601 format). The letter '*' indicates that the time zone is unknown.

# intake_height_end_date :	2016-06-01T00:00:00+09:00	20 or 25
# instrument_total_listed :	2	Variable
# instrument# :	HORIBA, Ltd. GA-360S	0-1000
# instrument#_measurement_method_type_code :	9	1 or 2
# instrument#_measurement_method_name :	NDIR	0-1000
# instrument#_start_date :	1994-01-01T00:00:00+09:00	20 or 25
# instrument#_end_date :	2009-12-31T23:59:59+09:00	20 or 25
# instrument# :	Round Science Inc. RGC-1	0-1000
# instrument#_measurement_method_type_code :	2	1 or 2
# instrument#_measurement_method_name :	Gas chromatography (FID)	0-1000
# instrument#_start_date :	2010-01-01T00:00:00+09:00	20 or 25
# instrument#_end_date :	2017-11-01T00:00:00+09:00	20 or 25
# scale_total_listed :	1	Variable
# scale#_code :	1	1-4
# scale#_name :	WMO CO2 X2007	1-100
# scale#_start_date :	1993-12-01T00:00:00+09:00	20 or 25
# scale#_end_date :	2016-06-01T00:00:00+09:00	20 or 25
# observation_status_code :	1	1
# observation_status :	Operational/Reporting	Variable
# measurement_calibration :	Carbon dioxide (CO2) in ambient and standard air samples is detected using a non-dispersive infrared (NDIR) analyzer. The measurement of CO2 in air is made relative to standards whose CO2 mole fraction is determined with high precision and accuracy. Because detector response is non-linear in the range of atmospheric levels, ambient samples are bracketed during analysis by a set of reference standards used to calibrate detector response.	0-3500
# sampling_frequency_code :	2	1 or 2
# sampling_frequency :	weekly	Variable
#		
# VARIABLE ATTRIBUTES		
#		
# site_gaw_id:long_name :	unique_station/mobile_identifiers.	34
# site_gaw_id:comment :	3-letter site identification code as defined by GAWGIS for stationary platforms	79
# time_components:FillValue :	-999 or -9	10
# time_components:long_name :	integer_components_of_LST_date/time	35
# time_components:order :	year, month, day, hour, minute, second	38
# time_components:comment :	Air sample collection time (LST) represented as a 6-element array [year, month, day, hour, minute, second]. Calendar time components as integers.	145
# value:FillValue :	-999.999	8
# value:units :	ppm	1-20
# value:long_name :	measured_mole_fraction_of_trace_gas_in_dry_air	0-51
# value:comment :	Reported mole fraction, isotope ratio or radioactivity. Units depend on trace gas species.	90
# value_unc:FillValue :	-999.999	8
# value_unc:units :	ppm	1-20
# value_unc:long_name :	estimated_uncertainty_in_reported_value	39
# value_unc:comment :	Standard deviation of the reported mean value when nvalue is greater than 1. Units depend on trace gas species.	111
# nvalue:FillValue :	-9	2
# nvalue:long_name :	number_of_measurements_contributing_to_reported_value	53
# nvalue:comment :	Number of individual measurements used to compute reported values.	66
# latitude:FillValue :	-999.999999999	14
# latitude:standard_name :	latitude	8
# latitude:long_name :	sample_latitude_in_decimal_degrees	34
# latitude:units :	degrees.north	13
# latitude:comment :	Latitude of sampling location in decimal degrees (north: +; south: -)	69
# longitude:FillValue :	-999.999999999	14
# longitude:standard_name :	longitude	9
# longitude:long_name :	sample_longitude_in_decimal_degrees	35
# longitude:units :	degrees.east	12
# longitude:comment :	Longitude of sampling location in decimal degrees (east: +; west: -)	68
# altitude:FillValue :	-999.999	8
# altitude:standard_name :	altitude	8
# altitude:long_name :	sample_altitude_in_meters_above_sea_level	41
# altitude:units :	m	1
# altitude:comment :	Altitude (elevation + intake height) of air sample collection. Units are meters above sea level (masl).	103
# elevation:FillValue :	-999.999	8
# elevation:standard_name :	elevation	9
# elevation:long_name :	surface_elevation_in_meters_above_sea_level	43
# elevation:units :	m	1
# elevation:comment :	Station height (m) above sea level	34
# intake_height:FillValue :	-999.999	8
# intake_height:long_name :	sample_intake_height_in_meters_above_ground_level	49
# intake_height:units :	m	1
# intake_height:comment :	Sampling height (depth) of air (seawater) above ground (below sea level) (height: +; depth: -) (m)	98
# flask_no:FillValue :	-999.999	8
# flask_no:long_name :	sample_flask_no	15
# flask_no:comment :	Identification code (or number) of flask used for observation	61
# ORG_QCflag:FillValue :	-999.999	8
# ORG_QCflag:long_name :	original_quality_control_flag	29
# ORG_QCflag:comment :	Details of original data quality control flags should be specified by contributors in metadata.	95
# QCflag:FillValue :	-9	2
# QCflag:long_name :	quality_control_flag	20
# QCflag:comment :	1 Valid (background) data considered "background" 2 Valid data considered valid other than "background" 3 Invalid data considered unfit for use (questionable/erroneous/absent data) Details of correspondence between original data quality control flags and WDCGG data flags should be specified by contributors in metadata.	49-139
# instrument:FillValue :	-9	2
# instrument:long_name :	instrument_ID_to_detect_atmospheric_parameter	45
# instrument:comment :	A numeric field that identifies the instrument.	47
# measurement_method:FillValue :	-9	2
# measurement_method:long_name :	measurement_method_history	26
# measurement_method:comment :	A numeric field that identifies the measurement method.	55
# scale:FillValue :	-9	2
# scale:long_name :	calibration_scale_history	25
# scale:comment :	A numeric field that identifies the calibration scale. Measurements are relative to reported calibration scale.	111
#		
# VARIABLE ORDER		
#		

Data of last item in data during the intake height period (ISO 8601 format). The letter '*' indicates that the time zone is unknown.
Number indicating how many kinds of instruments
Product name of instrument. The letter '#' represents a number from 1 to maximum (instrument_total_listed).
Measurement method identification number
-
Date of first item in data during the instrument period (ISO 8601 format). The letter '*' indicates that the time zone is unknown.
Data of last item in data during the instrument period (ISO 8601 format). The letter '*' indicates that the time zone is unknown.
Product name of instrument. The letter '#' represents a number from 1 to maximum (instrument_total_listed).
Measurement method identification number
-
Date of first item in data during the instrument period (ISO 8601 format). The letter '*' indicates that the time zone is unknown.
Data of last item in data during the instrument period (ISO 8601 format). The letter '*' indicates that the time zone is unknown.
Number indicating how many kinds of scales.
Scale identification number. The letter '#' represents a number from 1 to maximum (scale_total_listed).
scale name
Date of first item in data during the scale period (ISO 8601 format). The letter '*' indicates that the time zone is unknown.
Data of last item in data during the scale period (ISO 8601 format). The letter '*' indicates that the time zone is unknown.
Observation status identification number
Observation situation
Calibration for determining the mole fractions
Sampling frequency identification number
Sampling frequency of source data
site_gaw_id : Site gaw id is an abbreviation for the sampling site name.
time_components : Air sample collection time (LST or UTC) represented as a 6-element array [year, month, day, hour, minute, second]. Calendar time components as integers.
value : Reported mole fraction or isotope ratio. Units depend on trace gas species.
value_unc : Standard deviation of the reported mean value when nvalue is greater than 1. Units depend on trace gas species.
nvalue : Number of individual measurements used to compute reported values.
latitude : Latitude of sampling location in decimal degrees (north: +; south: -)
longitude : Longitude of sampling location in decimal degrees (east: +; west: -)
altitude : Altitude (elevation + intake height) of air sample collection. Units are meters above sea level (masl).
elevation : Station height (m) above sea level
intake_height : Sampling height (depth) of air (seawater) above ground (below sea level) (height: +; depth: -) (m)
flask_no : Identification code (or number) of flask used for observation
ORG_QCflag : Details of original data quality control flags should be specified by contributors in metadata.
QCflag : Quality control flag value that are delivered from data centers to users. (1:Data considered "background", 2:Data considered valid other than "background", 3:Data considered unfit for us (questionable/erroneous/absent data))
instrument : A numeric field that identifies the instrument.
measurement_method : A numeric field that identifies the measurement method.
scale : A numeric field that identifies the calibration scale. Measurements are relative to reported calibration scale.

site_gaw_id year month day hour minute second year month day hour minute second value value_unc nvalue latitude longitude altitude elevation intake_height flask_no ORG_Qcflag QCflag instrument measurement_method scale

2 Data record format

Data records in the new WDCGG file format are as follows (“+” represents a space):

[Site_gaw_id]+[Year]+[Month]+[Day]+[Hour]+[Minute]+[Second]+[Year]+[Month]+[Day]+[Hour]+[Minute]+[Second]+[Value]+[Value_unc]+[Nvalue]+[Latitude]+[Longitude]+[Altitude]+[Elevation]+[Intake_height]+[Flask_no]+[ORG_QCflag]+ [QC flag] + [Instrument] + [Measurement_method] + [Scale].

Example:

```
# site_gaw_id year month day hour minute second year month day hour minute second value
value_unc nvalue latitude longitude altitude elevation intake_height flask_no ORG_QCflag QCflag
instrument measurement_method scale
RYO 1987 01 01 00 00 00 -999 -9 -9 -9 -9 -9 353.15 0.959 227 39.033000946 141.8170013428 280 260 20
-999.999 3 2 1 9 1
```

Composition items are as per the table below.

Table. Data record elements

Item	Number of digits	"No Data" expression	Content	Detail
Site_gaw_id	3	-	Site code	3-letter site identification code as defined by GAWSIS for stationary platforms
Year	4	-999	Start year	Initial calendar year of observation
Month	2	-9	Start month	Initial calendar month of observation
Day	2	-9	Start day	Initial day of observation
Hour	2	-9	Start hour	Initial hour of observation
Minute	2	-9	Start minute	Initial minute of observation
Second	2	-9	Start second	Initial second of observation
Year	4	-999	End year	Final calendar year of observation
Month	2	-9	End month	Final calendar month of observation
Day	2	-9	End day	Final day of observation
Hour	2	-9	End hour	Final hour of observation
Minute	2	-9	End minute	Final minute of observation
Second	2	-9	End second	Final second of observation
Value	Variable	-999.999	Observation value	Reported mole fraction, isotope ratio or radioactivity. Units depend on trace gas species.
Value_unc	Variable	-999.999	Uncertainty based on standard deviation	Standard deviation of the reported mean value when Nvalue is greater than 1. Units depend on trace gas species.
Nvalue	Variable	-9	Number of measurements	Number of individual measurements used to compute reported values
Latitude	Variable	-999.999999999	Latitude	Latitude of sampling location in decimal degrees (north: +; south: -)
Longitude	Variable	-999.999999999	Longitude	Longitude of sampling location in decimal degrees (east: +; west: -)
Altitude	Variable	-999.999	Elevation + intake height	Altitude (elevation + intake height) of air sample collection. Units are meters above sea level (masl).
Elevation	Variable	-999.999	Station height	Station height (m) above sea level
Intake_height	Variable	-999.999	Sampling height	Sampling height (depth) of air (seawater) above ground (below sea level) (height: +; depth: -) (m)
Flask_no	Variable	-999.999	Flask identifier	Identification code (or number) of flask used for observation
ORG_QCflag	Variable	-999.999	Original data quality control flag	Details of original data quality control flags should be specified by contributors in metadata.
QCflag	Variable (1 or 2)	-9	WDCGG data flag	1 Valid (background) data considered "background" 2 Valid data considered valid other than "background" 3 Invalid data considered unfit for use (questionable/erroneous/absent data) Details of correspondence between original data quality control flags and WDCGG data flags should be specified by contributors in metadata.

Instrument	Variable (1 or 2)	-9	Instrument number	Instrument specification indicated in header
Measurement_method	Variable (1 or 2)	-9	Measurement method code	Measurement method employed.
Scale	Variable (1 to 4)	-9	Scale code	Employed scale in observation