

# WDCGG Contributor Manual

## Chapter 1 2018 WDCGG website update

### 1. 1 Objectives of the 2018 WDCGG website update

The World Data Centre for Greenhouse Gases (WDCGG) has been operating for more than a quarter of a century since its establishment in 1990. Requests from data contributors and users have included:

- a. Notification regarding data usage (e.g., user names/affiliations, purpose of data usage) to contributors
- b. Unification of archive data file formats in various categories to a single format for simplified processing on users' computer systems.

Certain types of observation data do not correspond fully with the format of metadata in the WDCGG database due to inappropriate metadata updates. As this makes it difficult for users and for WDCGG staff to improve database management and data services, WDCGG improved the database and strengthened links between observation data and metadata. Along with these improvements, WDCGG updated its official website and data submission procedures.

### 1.2 Data submission changes

In the previous WDCGG system, data were sent by email, metadata were updated via an online form by contributors, and there was no procedure for user identification. Now there is a registration page on the website for contributors and users.

Major changes regarding data submission on the WDCGG website:

- a. Contributors are strongly encouraged to submit both observation data and corresponding metadata via the website.
- b. Previously there was no particular metadata format. Now certain metadata elements (e.g., aims of observation) are submitted with ready-made descriptions.
- c. A single standard data file format for submission, which does not depend on observation platforms, is defined to facilitate final data usage.

## Chapter 2 Data submission

Before submitting data, contributors need to sign up for an account on the WDCGG website (<https://gaw.kishou.go.jp/>). Accounts are assigned to organizations rather than to individuals. Contributors log in, click “Data/Metadata Submission” and follow five simple steps to submit data as outlined below.

### 2.1 Step 1. Select Data Catalogue

Check the relevant gas species for each station from which data is to be submitted. At least one box must be checked.

### 2.2 Step 2. Input Metadata/Edit Metadata

To input or edit metadata, select the item, click “Edit” and fill out the form on each page. “Organization,” “Contact ,” “Aim of Observation,” “Data Time Zone” and “Unit” are required fields. Once metadata input or editing is complete, click “Next” at the bottom of the page.

### 2.3 Step 3. Check Metadata

Check the updated metadata carefully. To edit, click “Back” at the bottom of the page to return to Step 2; otherwise, click “Next.”

### 2.4 Step 4. Check Station

Check the station information from the GAW Station Information System (GAWSIS). Check and edit the mobile station information stored in the WDCGG database as needed. The station/mobile information here is shown in the header part of the WDCGG data file. Latitude, longitude and altitude values differing from GAWSIS values should be included in the data part of the file. Then, click “Next.”

Be sure to check the station information. WDCGG just obtains station information from the GAWSIS. Please update the station information on the GAWSIS website as needed.

### 2.5 Step 5. Upload Datafile

5 files(under <10MB each) are allowed to select. File Upload will start after clicking “Submit.” A single compressed file is highly encouraged.

Limit: 5 Files (less than 10 MB per file)

Allowed extensions: csv, txt, dat, xls, xlsx, prn, zip, tar.

Contributors should input the URL to have WDCGG retrieve files from their FTP server. Enter the name and email address on the form and click “Submit.”

Follow the online instructions for “Data/Metadata Submission.” WDCGG will contact contributors at the email address provided once data processing is complete.

## Chapter 3 Metadata Submission

Observation data and metadata should be updated in a timely manner to ensure that WDCGG data matches the latest version of submitters' data. Contributors are asked to submit details of stations, observations and other related information as metadata. As these data are indispensable for effective usage, descriptions should be as outlined on the WDCGG website. Required fields are marked with a red asterisk. Changes in metadata such as calibration scales and observation instruments can be specified (via submission) for the corresponding period. To delete metadata, leave the field blank and proceed to the next step.

### 3.1 Organization\*

The organization submitting the relevant data. The organization's name, acronym and country/territory are required. Enter the Organization Acronym field using upper and/or lower case letters, like "WDCGG" or "Empa," and 2 to 16 characters. The table on the page shows contributor information as registered on the WDCGG database. Please check it carefully, and modify it if needed.

### 3.2 Collaborator(s)

Institutes or organizations collaborating with the contributor. The organization's name, acronym and country/territory are required. The organization acronym should be entered using upper and/or lower case (e.g., WDCGG), and should be between 2 and 16 characters.

Up to three collaborators can be chosen from the drop-down list at one step.

### 3.3 Contact\*

The contact is the person responsible for handling inquiries, requests and other forms of consultation regarding submitted data. At least one person must be registered, and multiple contacts are preferable. WDCGG and data users can email the contact directly.

Contacts can be chosen from the drop-down list. If no list entries are applicable, add contact details as necessary (name, organization, role, country/territory and so on).

### 3.4 Aim of Observation\*

For "Aim of Observation," choose "Background observation," "Air pollution monitoring," "Biosphere monitoring" or "Atmospheric tracer." If none of these applies, specify the aim in the box that appears after the "Other" option is selected. Entering "Unknown" is discouraged.

### 3.5 Data Time Zone\*

Time zones are defined by the time offset (i.e., the difference between local time and UTC (local time – UTC)) from Coordinated Universal Time (UTC). Select the time zone for the timestamp of the observation data. If none of these applies, specify the relevant information after selecting the "Other" option. Entering "Unknown" is discouraged.

### 3.6 Unit\*

Select the unit used for the observation data. If none of these applies, specify the relevant information after selecting the "Other" option. Entering "Unknown" is discouraged.

### 3.7 Calibration Scale

The calibration scale used for observation data and the period of the scale's usage should be indicated here. If the period is equal to the entire period of the observation, input the date and time as follows:

Start date and time: 9999-12-31, T00:00:00

End date and time: 9999-12-31, T23:59:59

Use the time zone specified in 3.5. Indications of "unknown" are discouraged for CO<sub>2</sub>, CH<sub>4</sub> and N<sub>2</sub>O in particular.

### 3.8 Instrument(s)

Indicate the characteristics of the measurement instrument(s). The period of usage, manufacturer, product name and model number should be provided. If the metadata are equal to the entire period of observation, input the date and time as follows:

Start date and time: 9999-12-31, T00:00:00

End date and time: 9999-12-31, T23:59:59

Use the time zone specified in 3.5.

### 3.9 Intake Height above ground level

The height (depth) of air (seawater) sampling above the ground (below sea level) in meters: positive for height and negative for depth, except for mobiles (3D) whose altitudes are included in the observation data. Metadata being equal over the entire period of the observation, input the date and time as follows:

Start date and time: 9999-12-31, T00:00:00

End date and time: 9999-12-31, T23:59:59

Use the time zone specified in 3.5. For mobile platforms, leave the field blank or input "NA."

### 3.10 Sampling Frequency

Select a Sampling Frequency used in the current observation. If none of these applies, specify the relevant information after selecting the "Other" option. Entering "Unknown" is discouraged.

### 3.11 Measurement Calibration

Contributors should indicate the details of measurement calibration. Calibration for determining mole fractions is described here along with procedures for analysis such as the order of introduction (sequence) for sample and standard gases (or zero gases) to the instrument, the relevant time period and the number of calibration points.

#### Example 1 (CO<sub>2</sub>)

The quadratic fitting curve is determined every two hours based on a set of five working standard gases consisting of CO<sub>2</sub>-in-natural-air mixtures with different concentrations. The mole fraction is determined from this fitting curve.

#### Example 2 (CH<sub>4</sub>)

Ambient and standard air samples are injected into a gas chromatograph (GC) with a sampling valve. Methane (CH<sub>4</sub>) is separated from other sample constituents using packed columns and detected using flame ionization (FID). This process is highly automated for field and laboratory operations. The instrument response for the sample must be compared to a standard with known CH<sub>4</sub> content.

### 3.12 Data Processing

Details of instrumental data processing and averaging are described here. Criteria used in selection

for data processing should also be described.

#### Example

Raw data from the instrument are collected via the data acquisition system and stored as raw data for each 30-second period. CO<sub>2</sub> observations are checked on site by an operator every day.

### 3.13 Processing for averaging

Details of processes on hourly, daily, monthly data or data selections regarding quality are to be provided here. Individual specifications on how hourly, daily and monthly averages are calculated can be given by selecting “Hourly”, “Daily” or “Monthly” first. Be sure to select the appropriate radio button.

#### Example

Processing of hourly data: Data are taken from the arithmetic mean of observations covering a one-hour period after rejection of irregular data and obvious errors.

### 3.14 Original Data Quality Flags

Indicate the criteria for data flagging. If data are already marked with WDCGG data quality flags in the next item description, original data quality flags may not be necessary depending on metadata for the next Data Quality Flag (Valid/Invalid) item.

#### Example 1:

Flag: Data Category

0: Background data

1: Data possibly affected by pollution (wind direction is W – SW)

2: Insufficient number of averaging data

3: Invalid data

#### Example 2

NOAA ESRL uses a 3-column quality control flag where each column is defined as follows:

column 1 Rejection flag. An alphanumeric other than a period (.) in the FIRST column indicates a sample with obvious problems during collection or analysis. This measurement should not be interpreted.

column 2 Selection flag. An alphanumeric other than a period (.) in the SECOND column indicates a sample that is likely valid but does not meet selection criteria determined by the goals of a particular investigation.

column 3 Information flag. An alphanumeric other than a period (.) in the THIRD column provides additional information about the collection or analysis of the sample.

WARNING: A "P" in the 3rd column of the QC flag indicates the measurement result is preliminary and has not yet been carefully examined by the PI. The "P" flag is removed once the quality of the measurement has been assessed.

### 3.15 Data Quality Flag (Valid/Invalid)

WDCGG has introduced its own preliminary data flagging scheme to make data files machine-readable. There are three basic types: “1 valid as background”, “2 valid”, and “3 invalid”. Indication should be provided on how flags relate to these three types.

This description helps to clarify the relationship between contributors’ original data quality flags and WDCGG data quality flags, and eliminates the need to input “QCflag” item values in data records as described in 4.3.

### **3.16 Reference(s)**

References to the observation (such as the instruments, data processing, and calibration) in the literature or URLs should be described here.

### **3.17 Observation Status**

Select the most appropriate current description from “Operational/Reporting,” “Temporarily suspended” and “Terminated.” Updates should be provided as appropriate.

### **3.18 Description**

Any other information on observations or data can be provided here.

# Chapter 4 Data File Names and Record Formats

## 4.1 Data file names

WDCGG employs the following naming convention:

```
{ WDCGGParameterCode }_{ SiteGawId }_{ Platform }_{ SamplingType }_{ Organization Code }_{ DataType }.txt
```

File names consists of WDCGGParameterCode, SiteGawId, Platform, Sampling\_Type, OrganizationCode and DataType. These must be delimited with underscores (“\_”) and dashes (“-”).

Example

```
co2_mnm_surface-insitu_1_monthly.txt
```

This expresses:

```
co2 (WDCGGParameterCode)_mnm(SiteGawId)_surface(Platform)-insitu(Sampling Type)_1(Organization Code)_monthly(DataType).txt
```

### (1) WDCGGParameterCode

This code identifies an observing gas species.

Details: [https://gaw.kishou.go.jp/documents/db\\_list/gas\\_species](https://gaw.kishou.go.jp/documents/db_list/gas_species)

### (2) SiteGawId

Three-letter codes in lower case are used to identify stations. For GAW stations, GAWSIS station IDs are employed.

Details: [https://gaw.kishou.go.jp/documents/db\\_list/station](https://gaw.kishou.go.jp/documents/db_list/station)

### (3) Platform

This code identifies a platform at which observation is performed.

Details: [https://gaw.kishou.go.jp/documents/db\\_list/platform](https://gaw.kishou.go.jp/documents/db_list/platform)

### (4) SamplingType

This code identifies a sampling type.

Details: [https://gaw.kishou.go.jp/documents/db\\_list/sampling\\_type](https://gaw.kishou.go.jp/documents/db_list/sampling_type)

### (5) OrganizationCode

This number is used to identify a contributor’s affiliation.

Details: [https://gaw.kishou.go.jp/documents/db\\_list/organization](https://gaw.kishou.go.jp/documents/db_list/organization)

### (6) DataType

This field indicates the temporal resolution for each data record.

Details: [https://gaw.kishou.go.jp/documents/db\\_list/data\\_type](https://gaw.kishou.go.jp/documents/db_list/data_type)

## 4.2 Header record

Header records are created automatically from metadata registered by data contributors. Lengths are variable and defined at the top of the record. For more information, see [https://gaw.kishou.go.jp/static/documents/wdcgg\\_data\\_format\\_table.pdf](https://gaw.kishou.go.jp/static/documents/wdcgg_data_format_table.pdf)

## 4.3 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].
```

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
RYO 1987 01 01 00 00 00 -999 -9 -9 -9 -9 353.15 0.959 227 39.033000946 141.8170013428 280 260 20
-999.999 3 -9
```

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.