argo data management

Description of the Argo GDAC File Merge Process

Version 1.2
March 2017

ARGO
part of the integrated global observation strategy

Argo data management

Description of the Argo GDAC File Checks

Authors: Mark Ignaszewski / FNMOC

How to cite this document

Mark Ignaszewski / FNMOC, **Description of the Argo GDAC File Merge Process**.
http://doi.org/10.13155/52154

Table of contents

History of the document

Version	Date	Comment
1.0	January 2016	Original version sent around for comment
1.1	May 2016	Add section about variable attributes. Documented that HISTORY is not in the merge-file.
1.2	March 2017	Add data validation section: Define the data validation checks that are performed and the response when a check fails. Specify how DATE_CREATION and DATE_UPDATE are managed. Add global attributes. Add variable attributes.

1   Introduction

This document describes the process of creating "merge files" – a merged core-argo/bio-argo file. The core-file and bio-file are submitted to the GDACs by the DACs. The merge-file is created at the GDACs.

The term "merge-pair" will be used to refer to the corresponding core-argo/bio-argo pair of files.

In the following discussion, the dimensions of a variable are often referred to. For the purpose of these discussions, the "STRING" dimensions are ignored since they merely provide the space to hold a single string. For instance, the declaration of "STATION_PARAMETERS(N_PROF, N_PARAM, STRING16)" is considered to be dependent only on (N_PROF, N_PARAM).

In the examples below, the variable indices are referred to. These indices start at 1, not 0 (zero) as in many programming languages.

This highlighting is used to indicate areas that need further discussion / decisions.

1.1   Requirements Validation

There are requirements placed on the data contained within the merge-pair files. The FileChecker does not enforce these requirements. The merge-file processing will enforce these requirements.

See §2.14 (Merge-pair Data Validation) for the details.

2   Profile merge-file

2.1   Merge-file Naming

The merge-file naming template is: M*d*<*float-num*>_<*cyc*>.nc

where:

- *d* is the overall data-mode of the file and is the same as the first character of the core-file – either "R" or "D"
- <*float-num*> and <*cyc*> are the float-number and cycle number from the merge-pair

2.2   Dimensions

DATE_TIME and **STRING***x*: These dimensions are essentially constants. They are set to their typical values as documented in the User's Manual.

N_PROF: Set to the same value as N_PROF in both the core-file and bio-file, which are required to be the same. (See §2.14 for the validation details.)

N_PARAM: Described in the next section.

N_LEVELS: This setting is required to be the same in the the core-file and bio-file. N_LEVELS in the merge-file will be set to this value. (See §2.14 for the validation details.)

N_CALIB: The N_CALIB setting in the merge-file will be the maximum of the settings in the core-file and bio-file.

N_HISTORY: History variables are not included in the merge-file. This dimension is included in the merge file as the UNLIMITED dimension for symmetry with the core- and bio-files but its value will always be 0.

2.3 Physical Parameters and N_PARAM

The physical parameters are divided into three categories in the "Argo physical parameters list" on the ADMT website (reference table 3 in the User's Manual): the core parameters – category "c"; the bio parameters – category "b"; the intermediate parameters – category "i".

The merge-file will contain the physical parameters identified as core ("c") and bio ("b") parameters.

The N_PARAM dimension is the maximum number of parameters that will be in single profile of the merge-file. To determine this, both of the merge-pair files must be examined, profile-by-profile to determine which <PARAM>s will be transferred for each profile.

NOTE that the process of selecting which parameters to merge from each file, and profile therein, establishes the *template* for the data merging process discussed in §2.9, §2.10, and §2.11, below.

SPECIAL CASE: If there is a case where the only parameter in the core-file is PRES and the only parameters in the bio-file are PRES and i-parameters, the merge-file will include the profile with PRES being the only parameter, to maintain the one-to-one profile correspondence between the core-, bio-, and merge-files.

Example 1: Selecting merge <PARAM>s and N_PARAM. The core-file contains only core-parameters, by definition. The bio-file paramter-types are indicated following the parameter name.

Core-file N_PROF=2; N_PARAM=3	Bio-file N_PROF=2 ; N_PARAM=5	Merge-file N_PROF=2 ; **N_PARAM=4**
Profile 1 Number of parameters=3	Profile 1 Number of parameters=2	Profile 1 Number of Parameters=4
STATION_PARAMETERS 1,1) PRES (c) 1,2) TEMP (c) 1,3) PSAL (c)	STATION_PARAMETERS 1,1) PRES (c) 1,2) DOXY (b)	STATION_PARAMETERS 1,1) PRES 1,2) TEMP 1,3) PSAL 1,4) DOXY
Profile 2 Number of parameters=1	Profile 2 Number of parameters=5	Profile 2 Number of parameters=3
STATION_PARAMETERS 2,1) PRES (c)	STATION_PARAMETERS: 2,1) PRES(c) 2,2) BBP(b) 2,3) UV_INTENSITY_REF_NITRATE(i) 2,4) MOLAR_NITRATE(i) 2,5) NITRATE(b)	STATION_PARAMETERS 2,1) PRES 2,2) BBP 2,3) NITRATE

2.4 Variable Definitions

The variable definitions (type and dimensions) of the merge-file variables are the same as those for the **bio-file**. The core-file and bio-file definitions are *nearly* identical. However, there are a couple of differences. The bio-file definitions are used for the merge-file.

Parameter names dimension: Merge-file parameter names are dimensioned to 64 characters. Parameter names are dimensioned to 16 characters in core-files and 64 characters in bio-files.

Applicable variables: STATION_PARAMETERS, PARAMETER, HISTORY_PARAMETER.

DATA_TYPE dimension: Merge-file DATA_TYPE will be dimensioned to 32 characters. DATA_TYPE is dimensioned to 16 characters in core-files and 32 characters in bio-files.

PARAMETER_DATA_MODE variable: This variable is an OPTIONAL variable in the bio-file. Merge-files will always contain this variable. The filling of this variable is described below.

2.5 Variable Attributes

Variables that are in both core- and bio-file: The attribute settings from the core-file will be transferred to the merge-file, including any additional/optional attributes. (Additional attributes are specifically allowed.)

These variables include:

- "General information variables" (sections 2.6 and 2.7).
- STATION_PARAMETER (section 2.9)
- Calibration variables (section 2.11)

Note: The implication of this is that only the settings and additional attributes from the core-file appear in the merge-file.

<PARAM> variables: These variables only occur in one of the merge-pair files, *except for the special case of PRES*. The attribute settings for <PARAM> variables will be transferred to the merge-file from the relevent file (core- or bio-file).

Special case: PRES occurs in both the core- and bio-file. Only the attributes from the core-file will be transferred to the merge-file.

2.6 "General informaton on the file" Variables

Variables described in the Argo User's Manual, §2.2.3.

These variables are characteristics of the merge file and are not merged from the input merge-pair files. Includes: DATA_TYPE, FORMAT_VERSION, HANDBOOK_VERSION, REFERENCE_DATE_TIME, DATE_CREATION, DATE_UPDATE.

*** NOTE: The following times are set by the GDAC when the files are created and/or updated. For a given file, they will vary between the two GDACs based on each GDAC's processing schedule.

DATE_CREATION: The time the file is first created at the GDAC. When the file is updated, this time is maintained.

DATE_UPDATE: The time the file was updated at the GDAC.

2.7 "General information for each profile" Variables

Variables described in the Argo User's Manual, §2.2.4 (except as noted below).

These variables describe information about each individual profile in the merge-pair. A single value must be chosen that is representative of the merge-pair (with the exceptions that follow):

The following algorithm is used:

- If the core-file value is set (not FillValue) for a profile, transfer it to the merge-file

- If the core-file value is FillValue, transfer the bio-file value to the merge-file (even if it is FillValue)

Implications:

- DATA_MODE: If set in the core-file, the core-file DATA_MODE becomes the overall DATA_MODE for the profile and the setting in the bio-file is ignored (on a profile-by-profile basis). PARAMETER_DATA_MODE should be set to retain parameter-dependent data-mode differences.
- PROJECT_NAME, PI_NAME: Only one for each profile can be stored in the merge file.

Exceptions:

- STATION_PARAMETERS: The merging of this variable is described in §2.9 (STATION_PARAMETERS and PARAMETER_DATA_MODE).

2.8 PROFILE_<PARAM>_QC: This variable is *not* handled as described here. It is part of the physical parameter merge process described in §2.10 (STATION_PARAMETERS and PARAMETER_DATA_MODE

STATION_PARAMETERS is a "special case" variable from Argo User's Manual, §2.2.4. PARAMETER_DATA_MODE is an optional variable for the bio-file described in Argo User's Manual, §2.6.5.

There are two variables that are dependent on only the (N_PROF, N_PARAM) dimensions: STATION_PARAMETERS, PARAMETER_DATA_MODE. (As previously noted: Ignoring the "STRING" dimension since it merely provides space for the strings.)

The process of selecting the physical parameters to be included in the merge and determining N_PARAM is described in §2.2 and Example 1 is relevant here.

The merge-file STATION_PARAMETERS will be built for each profile as follows:

- the parameters (non-blank values) from the core-file are added to the merge-file STATION_PARAMETERS, in the same order as in the core-file
- then the bio-parameters from the bio-file are added to the merge-file STATION_PARAMETERS, in the same order as in the bio-file

Notes:

- The bio-file PRES is ignored in this process
- No blank values will be inserted with the sequence of valid parameter names

The PARAMETER_DATA_MODE will be built such that the data-mode associated with each parameter put into the merge-file STATION_PARAMETERS is placed into the same position in the merge-file PARAMETER_DATA_MODE.

Example 2: Merging STATION_PARAMETERS and PARAMETER_DATA_MODE. The paramter-types are indicated following the parameter name.

Core-file N_PROF=2; N_PARAM=3	Bio-file N_PROF=2 ; N_PARAM=5	Merge-file N_PROF=2 ; **N_PARAM=4**
Profile 1 Number of parameters=3	Profile 1 Number of parameters=2	Profile 1 Number of Parameters=4

STATION_PARAMETERS: 1,1) PRES (c) 1,2) TEMP (c) 1,3) PSAL (c)	STATION_PARAMETERS: 1,1) PRES (c) 1,2) DOXY (b)	STATION_PARAMETERS: 1,1) PRES 1,2) TEMP 1,3) PSAL 1,4) DOXY
PARAMETER_DATA_MODE 1,1) D 1,2) D 1,3) D	PARAMETER_DATA_MODE 1,1) R 1,2) A	PARAMETER_DATA_MODE 1,1) D 1,2) D 1,3) D 1,4) A
Profile 2 Number of parameters=1	Profile 2 Number of parameters=5	Profile 2 Number of parameters=3
STATION_PARAMETERS: 2,1) PRES (c)	STATION_PARAMETERS: 2,1) PRES(c) 2,2) BBP(b) 2,3) UV_INTENSITY_REF_NITRATE(i) 2,4) MOLAR_NITRATE(i) 2,5) NITRATE(b)	STATION_PARAMETERS: 2,1) PRES 2,2) BBP 2,3) NITRATE
PARAMETER_DATA_MODE 2,1) R	PARAMETER_DATA_MODE 2,1) R 2,2) R 2,3) R 2,4) R 2,5) A	PARAMETER_DATA_MODE 2,1) R 2,2) R 2,3) A

- <PARAM>).

2.9 STATION_PARAMETERS and PARAMETER_DATA_MODE

STATION_PARAMETERS is a "special case" variable from Argo User's Manual, §2.2.4.
PARAMETER_DATA_MODE is an optional variable for the bio-file described in Argo User's
Manual, §2.6.5.

There are two variables that are dependent on only the (N_PROF, N_PARAM) dimensions:
STATION_PARAMETERS, PARAMETER_DATA_MODE. (As previously noted: Ignoring the
"STRING" dimension since it merely provides space for the strings.)

The process of selecting the physical parameters to be included in the merge and determining
N_PARAM is described in §2.2 and Example 1 is relevant here.

The merge-file STATION_PARAMETERS will be built for each profile as follows:

- the parameters (non-blank values) from the core-file are added to the merge-file
  STATION_PARAMETERS, in the same order as in the core-file
- then the bio-parameters from the bio-file are added to the merge-file
  STATION_PARAMETERS, in the same order as in the bio-file

Notes:

- The bio-file PRES is ignored in this process
- No blank values will be inserted with the sequence of valid parameter names

The PARAMETER_DATA_MODE will be built such that the data-mode associated with each
parameter put into the merge-file STATION_PARAMETERS is placed into the same position in the
merge-file PARAMETER_DATA_MODE.

Example 2: Merging STATION_PARAMETERS and PARAMETER_DATA_MODE. The paramter-types are indicated following the parameter name.

Core-file N_PROF=2; N_PARAM=3	Bio-file N_PROF=2 ; N_PARAM=5	Merge-file N_PROF=2 ; **N_PARAM=4**
Profile 1 Number of parameters=3	Profile 1 Number of parameters=2	Profile 1 Number of Parameters=4
STATION_PARAMETERS: 1,4) PRES (c) 1,5) TEMP (c) 1,6) PSAL (c)	STATION_PARAMETERS: 1,3) PRES (c) 1,4) DOXY (b)	STATION_PARAMETERS: 1,5) PRES 1,6) TEMP 1,7) PSAL 1,8) DOXY
PARAMETER_DATA_MODE 1,4) D 1,5) D 1,6) D	PARAMETER_DATA_MODE 1,3) R 1,4) A	PARAMETER_DATA_MODE 1,5) D 1,6) D 1,7) D 1,8) A
Profile 2 Number of parameters=1	Profile 2 Number of parameters=5	Profile 2 Number of parameters=3
STATION_PARAMETERS: 2,2) PRES (c)	STATION_PARAMETERS: 2,6) PRES(c) 2,7) BBP(b) 2,8) UV_INTENSITY_REF_NITRATE(i) 2,9) MOLAR_NITRATE(i) 2,10)　　NITRATE(b)	STATION_PARAMETERS: 2,4) PRES 2,5) BBP 2,6) NITRATE
PARAMETER_DATA_MODE 2,2) R	PARAMETER_DATA_MODE 2,6) R 2,7) R 2,8) R 2,9) R 2,10)　　A	PARAMETER_DATA_MODE 2,4) R 2,5) R 2,6) A

2.10 <PARAM> Variables

The core-parameters and bio-parameters are included in the merge-file as described in §2.2; intermediate parameters are not included in the merge-file. The process of "merging" the <PARAM> variables simply means copying the variables associated with the merged parameters from the either the core-file or bio-file into the merge-file.

The <PARAM> variables to be transferred to the merge-file are: PROFILE_<PARAM>_QC, <PARAM>, <PARAM>_QC, <PARAM>_ADJUSTED, <PARAM>_ADJUSTED_QC, and <PARAM>_ADJUSTED_ERROR.

NOTE on PRES: PRES is required to be present in both the core- and bio-file; only PRES – no associated variables – are in the bio-file. The PRES variables (including _QC, etc) from the core-file are transferred into the merge-file. See §2.14 for the for the validation checks that are applied to PRES.

2.11 Calibration Variables

The calibration variables are those variable that have an N_CALIB dimension. They are:

- PARAMETER
- SCIENTIFIC_CALIB_EQUATION
- SCIENTIFIC_CALIB_COEFFICIENT
- SCIENTIFIC_CALIB_COMMENT
- SCIENTIFIC_CALIB_COMMENT.

All of these variables are dimensioned as (N_PROF, N_CALIB, N_PARAM, string-size). (In this discussion the "string" dimensions are going to be ignored.)

PARAMETER:

For each profile, fill the PARAMETER variable for all N_CALIB indices to the merge-file STATION_PARAMETERS for that profile.

SCIENTIFIC_*_*:

For each profile index:
- For each calibration index:
  - Transfer the calibration information from the core-file and bio-file into the same profile index and calibration index in the merge-file, but for the associated merge-file parameter index.

Example 3: Merging calibration variables. The core-file contains only core-parameters, by definition. The bio-file paramter-types are indicated following the parameter name.

Core-file N_PROF=2; N_CALIB=2 N_PARAM=3	Bio-file N_PROF=2 ; N_CALIB=3 N_PARAM=5	Merge-file N_PROF=2 ; N_CALIB=3 N_PARAM=4
Profile 1 Number of calibrations=2 Number of parameters=3	Profile 1 Number of calibrations=3 Number of parameters=2	Profile 1 Number of Parameters=4
STATION_PARAMETERS: 1,1) PRES (c) 1,2) TEMP (c) 1,3) PSAL (c)	STATION_PARAMETERS: 1,1) PRES (c) 1,2) DOXY (b)	STATION_PARAMETERS: 1,1) PRES 1,2) TEMP 1,3) PSAL 1,4) DOXY
PARAMETER: 1,1) PRES (c) 1,2) TEMP (c) 1,3) PSAL (c)	PARAMETER: 1,1) PRES (c) 1,2) DOXY (b)	PARAMETER: 1,1) PRES 1,2) TEMP 1,3) PSAL 1,4) DOXY
SCIENTIFIC_CALIB_* 1,1,1) PRES calib info #1 1,1,2) TEMP calib info 1,1,3) PSAL calib info #1 1,2,1) \<blank> 1,2,2) \<blank> 1,2,3) PSAL calib info #2	SCIENTIFIC_CALIB_* 1,1,1) PRES calib info #1 1,1,2) DOXY calib info #1 1,1,3) \<blank> 1,1,4) \<blank> 1,1,5) \<blank> 1,2,1) PRES calib info 1,2,2) DOXY calib info #2 1,2,3) \<blank> 1,2,4) \<blank> 1,2,5) \<blank> 1,3,1) \<blank> 1,3,2) DOXY calib info #3 1,3,3) \<blank> 1,3,4) \<blank> 1,3,5) \<blank>	SCIENTIFIC_CALIB_* 1,1,1) PRES calib info #1 1,1,2) TEMP calib info 1,1,3) PSAL calib info #1 1,1,4) DOXY calib info #1 1,2,1) \<blank> 1,2,2) \<blank> 1,2,3) PSAL calib info #2 1,2,4) DOXY calib info #2 1,3,1) \<blank> 1,3,2) \<blank> 1,3,3) \<blank> 1,3,4) DOXY calib info #3
Profile 2 Number of parameters=1 Number of calibrations=1	Profile 2 Number of parameters=5 Number of calibrations=1	Profile 2 Number of parameters=3
STATION_PARAMETERS: 2,1) PRES (c)	STATION_PARAMETERS: 2,1) PRES(c) 2,2) BBP(b) 2,3) UV_INTENSITY_REF_NITRATE(i) 2,4) MOLAR_NITRATE(i) 2,5) NITRATE(b)	STATION_PARAMETERS: 2,1) PRES 2,2) BBP 2,3) NITRATE

PARAMETER:	PARAMETER:	PARAMETER:
2,1) PRES (c)	2,1) PRES(c)	2,1) PRES
	2,2) BBP(b)	2,2) BBP
	2,3) UV_INTENSITY_REF_NITRATE(i)	2,3) NITRATE
	2,4) MOLAR_NITRATE(i)	
	2,5) NITRATE(b)	
SCIENTIFIC_CALIB_*	SCIENTIFIC_CALIB_*	SCIENTIFIC_CALIB_*
2,1,1) PRES calib info #2	2,1,1) PRES calib info #2	2,1,1) PRES calib info #2
2,1,2) <blank>	2,1,2) BBP calib info #1	2,1,2) BBP calib info #1
2,1,3) <blank>	2,1,3) UV_INT_REF calib info #1	2,1,3) NITRATE calib info #1
	2,1,4) MOLAR_NIT calib info #1	
2,2,1) <blank>	2,1,5) NITRATE calib info #1	
2,2,2) <blank>	2,2,1) <blank>	2,2,1) <blank>
2,2,3) <blank>	2,2,2) BBP calib info #2	2,2,2) BBP calib info #2
	2,2,3) UV_INT_REF calib info #2	2,2,3) NITRATE calib info #2
	2,2,4) MOLAR_NIT calib info #2	
	2,2,5) NITRATE calib info #2	
	2,3,1) <blank>	2,3,1) <blank>
	2,3,2) <blank>	2,3,2) <blank>
	2,3,3) <blank>	2,3,3) NITRATE calib info #3
	2,3,4) <blank>	
	2,3,5) NITRATE calib info #3	

2.12 History Variables

In many cases, the HISTORY variables are only relevent in the context of the file in which the occur (the core- or bio- file). Plus, merging the histories from both files could significantly increase the size of the merge-file. Therefore, HISTORY variables are *not* included in the merge-file.

2.13 Global Attributes

The global attributes are copied from the core-file. The global attributes from the bio-file are ignored.

Exception: The global attribute for "institution" is either "US GDAC" or "French GDAC".

2.14 Merge-pair Data Validation

There are requirements placed on the data contained within the merge-pair files. The FileChecker does not enforce these requirements. The merge-file processing will enforce these requirements.

In case of failures: If any of the following checks fail

- A warning is issued to the DAC
- No merge-file is created
- An existing merge-file is removed.
- The core-file and bio-file remain on the GDAC.

Merge-pair Validation Checks:

- FORMAT_VERSION: The merge-pair files must have the same FORMAT_VERSION
  - This requirement is currently being "relaxed" during the v3.0 to v3.1 transition
- N_PROF: The merge-pair files must have the same N_PROF value
- N_LEVELS: The merge-pair files must have the same N_LEVELS value
- PRES: On a profile-by-profile basis
  - The PRES level values must be the same between the core-file and bio-file

3 Trajectory merge file

The User's Manual refers to trajectory merge files. This process will be defined at a later time and is not currently implemented.