Good procedures to manage OceanGliders data with US glider DAC.

V1.3 – January 2020

Author:

NOTE: This document is a first introduction to US Gliders DAC data management. The complete description of the good practices is available here: <https://gliders.ioos.us/data/>

Following a **thorough** reading of US glider DAC guidelines, all additional questions, concerns or suggestions should be directed to: ioos.glider.data.providers@noaa.gov

Contents

[General information 2](#_Toc31026437)

[About OceanGliders 2](#_Toc31026438)

[About this document 2](#_Toc31026439)

[OceanGliders data management structure 2](#_Toc31026440)

[Data mode 2](#_Toc31026441)

[Format 2](#_Toc31026442)

[Data policy 3](#_Toc31026443)

[The good procedures for Data Provider 4](#_Toc31026444)

[Data Provider Registration 4](#_Toc31026445)

[New Deployment Registration 4](#_Toc31026446)

[Requesting a WMO ID 4](#_Toc31026447)

[Deployment Creation 5](#_Toc31026448)

[Submission of NetCDF Files 5](#_Toc31026449)

[Dataset Status 6](#_Toc31026450)

[Dataset Archiving 6](#_Toc31026451)

# General information

## About OceanGliders

OceanGliders is the international glider program of the GOOS. It is dedicated to the promotion of the glider technology for sustained Ocean observing and scientific studies. The OceanGliders community promotes glider applications through coordination, training, liaison between providers and users, advocacy, and provision of expert advice. We intend to favour oceanographic experiments and operational monitoring of the oceans with gliders through scientific and international collaboration. We provide information about glider projects and glider data management, as well as resources related to gliders. All OceanGliders data are publicly available.

## About this document

This document introduces to the good procedures for gliders data management for US Glider DAC users.

This document relies on the complete guideline available here:  <https://gliders.ioos.us/data/>

OceanGliders identified 2 GDACs, Coriolis and US Glider DAC. For the glider groups working with the Coriolis glider DAC, please refer to: <https://www.ego-network.org/dokuwiki/lib/exe/fetch.php?media=public:datamanagement:v1.3:good_practices_for_glider_data_management.docx>

## OceanGliders data management structure

Data flow within OceanGliders is organized at three stages: operator/PI, DAC and GDAC.

* Operator are deploying, piloting and recovering the gliders. They are responsible for the operation at sea and deliver the data and metadata in real time.
* The Principal Investigator (PI), typically a scientist, is responsible for the mission.
* The DAC processes raw data and metadata into Netcdf files following OceanGliders requirements DAC should deliver the EGO files to the GDAC in near real time.
* The GDAC archives and distributes the best copy of the data set. When a higher quality data set (e.g. calibrated data) is available, it replaces the previous version of the data file.

In some cases, PI and DAC as well as DAC and GDAC can be the same person or institution.

## Data mode

There is three temporal mode for gliders data: Real-Time data, High Frequency data (for Slocums) and delayed mode data.

* The Real-Time mode (RT) concerns the data that are transmitted by the glider when operating.
* The High-Resolution mode (HR), or recovery mode, is the full dataset collected after glider recovery. It is High Frequency and full parameter.
* The delayed mode (DM) are quality controlled data. The delayed mode data management procedures are not established yet.

## Format

The format processed by US Glider DAC is described here: <https://ioos.github.io/ioosngdac/ngdac-netcdf-file-format-version-2.html>

## Data policy

Real Time data in US Glider DAC are publicly available through public ERDDAP and THREDDS.

# The good procedures for US National GDAC (NGDAC) Data Provider

## Data Provider Registration

You must register as a data provider and receive a user account in order to contribute data sets to the **IOOS Glider Data Assembly Center (NGDAC)**. The following contact information is required:

* Contact name
* Contact organization
* Email address
* Telephone number

and all user account requests should be emailed to: glider.dac.support@noaa.gov

**New user accounts are typically created the same day they are received**.

## New Deployment Registration

Data providers can register deployments after the user account has been created. New deployment registration is a 2 steps process.

Step 1: you must request a [WMO id](http://www.wmo.int/pages/prog/amp/mmop/buoy-ids.html) for the glider. This ID must be referenced as a [global](https://github.com/ioos/ioosngdac/wiki/NGDAC-NetCDF-File-Format-Version-2#description--examples-of-required-global-attributes) attribute as well as an attribute of the file's [platform](https://github.com/ioos/ioosngdac/wiki/NGDAC-NetCDF-File-Format-Version-2#platform) variable in each NetCDF file uploaded to the **NGDAC**.

Step 2: register the deployment with the **NGDAC**.

### Requesting a WMO ID

In order for the datasets to be released to the [Global Telecommunication System](http://www.wmo.int/pages/prog/www/TEM/index_en.html), the glider must be assigned a **WMO ID**. All **WMO ID** requests should be sent to:

         ioos.glider.data@noaa.gov – or support@jcommops.org

The following information must be provided for each request:

* Data provider
* Glider name
* Glider model and Serial Number
* Approximate deployment date
* Approximate deployment location (GPS coordinates)

Once the request is received, it will be forwarded to the [National Data Buoy Center](http://www.ndbc.noaa.gov/) and the assigned **WMO ID** will be sent to the requestor/data provider.

**IMPORTANT**: WMO ids are assigned based on the [WMO region](http://en.wikipedia.org/wiki/Location_identifier#WMO_station_identifiers) in which the glider is deployed. Once assigned, the WMO id may be used on successive deployments of the same glider provided it is deployed in the same WMO region.

**You do not need to acquire a different WMO id each time the glider is deployed, provided it's deployed in the same WMO region.**

### Deployment Creation

Deployments are registered and managed via a [web page](http://data.ioos.us/gliders/providers). Each deployment must be registered by the data provider **before** any NetCDF files are uploaded. The deployment registration process is as follows:

1. Log in using your data provider credentials.
2. Click the **Your Deployments** link. A deployment registration form will be displayed.
3. Specify a unique deployment name using the following convention: **GLIDER-YYYYmmddTHHMM**

Where **GLIDER** is the actual name of the glider and **YYYYmmddTHHMM** is the timestamp specifying the start of the deployment. This is also the value that should be assigned to the [trajectory](https://github.com/ioos/ioosngdac/wiki/NGDAC-NetCDF-File-Format-Version-2#trajectory) variable in each NetCDF file that is submitted to the DAC. Enter the **WMO id** assigned to this glider for the new deployment in the form element.

**IMPORTANT**: This WMO id must also be referenced as a [global](https://github.com/ioos/ioosngdac/wiki/NGDAC-NetCDF-File-Format-Version-2#description--examples-of-required-global-attributes) attribute as well as an attribute of the file's [platform](https://github.com/ioos/ioosngdac/wiki/NGDAC-NetCDF-File-Format-Version-2#platform) variable in each NetCDF file uploaded to the **NGDAC**.

1. Click **New Deployment** to create the deployment. This creates a directory on the IOOS Glider DAC FTP server using the specified deployment name. This is the directory that the NetCDF files must be uploaded to.

**IMPORTANT: New deployments cannot be created by logging into the ftp server. All new deployments must be created via the process described above.**

1. After the deployment has been registered, click on the deployment name to take you to the deployment metadata page and specify the **operator**. Once the deployment has been completed (i.e.: the glider has been recovered or the deployment has been completed), click the **Completed** check box to denote that the data is ready for archiving by [NODC](http://www.nodc.noaa.gov).

## Submission of NetCDF Files

The data provider user account provides ftp push access to the directories created under the user's home directory. The ftp URL is:

         ftp://data.ioos.us

Here's an example of the ftp login process and the resulting directory structure:

|  |
| --- |
| $ ftp -i data.ioos.usConnected to data.ioos.us (54.204.42.247).220 Welcome to the IOOS Glider DAC FTP ServerName (data.ioos.us:kerfoot): rutgers331 Please specify the password.Password:230 Login successful.Remote system type is UNIX.Using binary mode to transfer files.ftp> pwd257 "/"ftp> dir200 PORT command successful. Consider using PASV.150 Here comes the directory listing.drwxr-xr-x 2 ftp ftp 77824 Dec 08 16:40 ru01-20140104T1621drwxr-xr-x 2 ftp ftp 4096 Dec 08 16:41 ru01-20140120T1444drwxr-xr-x 2 ftp ftp 57344 Dec 08 16:41 ru01-20140123T1250drwxr-xr-x 2 ftp ftp 65536 Dec 08 16:41 ru01-20140217T1244drwxr-xr-x 2 ftp ftp 36864 Jan 15 16:07 ru05-20150105T1600drwxr-xr-x 2 ftp ftp 65536 Feb 02 13:41 ru05-20150115T1443226 Directory send OK. |

New NetCDF files should be uploaded to the directory [created above](https://github.com/ioos/ioosngdac/blob/gh-pages/_docs/ngdac-netcdf-file-submission-process.md#deployment-creation). For example, uploading files to the ru05-20150115T1443 deployment is done as follows:

|  |
| --- |
|  ftp> cd ru05-20150115T1443 250 Directory successfully changed. ftp> lcd LOCAL\_DIRECTORY Local directory now LOCAL\_DIRECTORY ftp> mput \*.nc |

Please remember to use the [proper](https://github.com/ioos/ioosngdac/wiki/NGDAC-NetCDF-File-Format-Version-2#file-naming-conventions) file naming convention.

The resulting deployment directory structure will look something like this:

|  |
| --- |
|  /ru05-20150115T1443 profile1.nc profile2.nc profile3.nc ... |

A generic [ftp script](https://github.com/ioos/ioosngdac/blob/master/util/ncFtp2ngdac.pl), written in [Perl](http://www.perl.org/) is contained in the repository and may be used to upload the files to the **NGDAC**. The script requires the following Perl non-core modules:

* [Readonly](http://search.cpan.org/~roode/Readonly-1.03/Readonly.pm)
* [Net::FTP](http://search.cpan.org/~shay/libnet-1.25/Net/FTP.pm)

**You must specify your credentials in the $USER and $PASS variables contained in the script**.

## Dataset Status

Once one or more files have been successfully uploaded for the specified deployment, the [aggregation](https://github.com/ioos/ioosngdac/wiki/NGDAC-Architecture#data-assembly-center-architecture) process begins. As there are multiple file syncing and aggregation processes going on, it will take some time for the data access end points on both the [ERDDAP](http://data.ioos.us/gliders/erddap/tabledap/index.html) and [THREDDS](http://data.ioos.us/gliders/thredds/catalog.html) servers to be created and populated. The end-to-end processing pathway **currently takes 1 - 2 hrs**. We are actively working on ways to decrease this time frame.

We've built a [dataset status](http://data.ioos.us/gliders/status/) page to provide administrators and users with the ability to track datasets through the end-to-end process. The [home page](http://data.ioos.us/gliders/status/) displays a list of all data sets for which either/both the [ERDDAP](http://data.ioos.us/gliders/erddap/tabledap/index.html) and [THREDDS](http://data.ioos.us/gliders/thredds/catalog.html) are not yet available. Please check this page before emailing the DAC administrators regarding data set availability.

## Dataset Archiving

Once a glider deployment is **completed**, the dataset is added to the [National Ocean Data Center's](http://www.nodc.noaa.gov/) national ocean archive. **It is the responsibility of the individual data provider to mark the deployment as completed before it will be archived NODC**. A deployment is marked as complete by checking the **Completed** checkbox on the individual deployment creation page. Once this box has been checked, an [md5 checksum](http://en.wikipedia.org/wiki/MD5) is generated on the trajectory NetCDF file and the dataset is added to the archive.

**The details of the NODC archiving process are still under development, so datasets are not currently archived by NODC. We are working with NODC on this**.