U.S. IOOS National Glider Data Assembly Center – version 2.0

Description of the Architecture for the End-to-end Data Pathway

John Kerfoot

July 14, 2014

1. A new data provider contacts DAC administrators and requests a data provider account.
2. For each new glider deployment, the data provider requests a WMO ID from DAC administrators. The request is forwarded to NDBC to acquire the WMO ID.
3. The data provider registers the new deployment (<http://gliders.ioos.us>) by entering data about the deployment and glider:
   1. Creates **$DEPLOYMENT** name: **glider-YYYYmmddTHHMM**
   2. WMO ID
   3. Operator
   4. Estimated Deployment Date
   5. Estimated Deployment Location
4. Once the new **$DEPLOYMENT** has been created by the data provider, the DAC administrator will create the **$DEPLOYMENT** ftp directory location (<ftp://ftp.gliders.ioos.us>)

**$FTP\_ROOT/$USER/upload/$DEPLOYMENT**

1. The first time the data provider uploads one or more .nc files to the **$DEPLOYMENT** location, the DAC administrator will:
   1. Create an rsync location on the private ERDDAP:

**$PRIVATE\_ERDDAP\_ROOT/sync/$USER/$DEPLOYMENT**

* 1. Create the <dataset></dataset> XML in the private ERDDAP datasets.xml file
  2. Create the public ERDDAP wget destination:

**$PUBLIC\_ERDDAP\_ROOT/$USER/$DEPLOYMENT**

* 1. Create the ERDDAP <dataset></dataset> XML in the public ERDDAP datasets.xml file
  2. Create the THREDDS wget destination:

**$TDS\_ROOT/$USER/$DEPLOYMENT**

* 1. Create the THREDDS catalog.xml for the dataset
  2. Notify, by email, all subscribers (i.e.: DAC administrators, NDBC, NODC, etc.) about the new dataset: ERDDAP does not provide this service, so we will need to roll our own solution.

1. The DAC administrator will create a cronjob to rsync the ftp site to the private ERDDAP locations (task #**5a**).
2. The DAC administrator will script a wget call to download the aggregated deployment dataset from the private ERDDAP to both the public ERDDAP deployment destination (task #**5c**) and the THREDDS deployment destination (task #**5e**). The wget call will create a CF-compliant Contiguous Ragged Array (.ncCF) NetCDF file for the full deployment.
3. The private ERDDAP provides an email subscription service to notify end-users when the dataset has changed (i.e.: new files have been uploaded):
   1. DAC administrators
4. Upon receiving email notification that one or more data sets have changed (i.e.: new files have been added) on the private ERDDAP server, the DAC uses the wget script (task **#7**) to overwrite the existing .ncCF file with the updated version from the private ERDDAP server and ping the flag URL for the dataset on the public ERDDAP to notify the public ERDDAP of the change(s).
5. The public ERDDAP provides an email subscription service to notify end-users when the dataset has changed (i.e.: new files have been uploaded):
   1. DAC administrators
   2. NDBC
   3. The data provider as confirmation that the new files have been successfully uploaded.
   4. Others?
6. Once NDBC is notified (task **#10**) that one or more data sets have changed, they can retrieve the new data from either the public ERDDAP or THREDDS, perform NDBC QC, encode to BUFR and release to GTS.