

## VIII. Data Archive Procedures

PacIOOS, as a Federally funded program, is required to submit data to a national archive center. At present, PacIOOS is working with the National Centers for Environmental Information (NCEI) in Silver Spring<sup>42</sup> to assist in this regard. The NCEI regional representative, Pat Caldwell, has advised PacIOOS on the data submission forms and all necessary procedures.

More recently PacIOOS was able to establish an archiving process using the ERDDAP server. On a regular basis, data are aggregated over a fixed time period, put into the archive package format (with checksums) and put on the PacIOOS web accessible folder (WAF; <http://www.pacioos.hawaii.edu/archive/>). NCEI regularly scans this folder and pulls the archive packages.

The initial agreement was to have annual aggregate files generated just after the new year. This would allow PacIOOS an opportunity to conduct any delayed-mode processing (*e.g.*, QC-ing against the annual cycle). NCEI has suggested a more frequent update, *i.e.*, monthly, so this may change.

PacIOOS wanted to ensure that data archived at NCEI were final in all respects. Consequently, these archive packages were delayed until the variable formats were finalized. At issue were the representation of the QC flags, *e.g.*, what to call them, how to represent them (bits, floats or integers), and how they would be applied. The WQB data have been processed and put into this format. The near-shore sensor data are awaiting a major upgrade to DataTurbine, and should be ready in Fall 2020.

At present, NCEI does not generally archive model output or data-derived products. Additionally, PacIOOS serves many data sets that already have an archive mechanism in place. This includes wave data that are processed via CDIP, animal (shark) track data that will be sent to the IOOS function DAC (Animal Tracking Network, ATN), ACO data that are archived at NODC directly by the PI's, *etc.* Thus the data that PacIOOS will be submitting for long-term archive are limited to the following:

1. Near-shore sensors
2. Water quality buoys
3. High-frequency radar
4. Gliders

---

<sup>42</sup> Formerly the National Ocean Data Center (NODC)

The archive process is done via a Submission Information Form (SIF; [https://www.nodc.noaa.gov/media/pdf/submit/NODC\\_SubmissionInfoForm\\_v1.3.pdf](https://www.nodc.noaa.gov/media/pdf/submit/NODC_SubmissionInfoForm_v1.3.pdf)). The submission agreement includes the following:

1. list of parameters/observations being collected.
2. processing steps/quality control including final format.
3. timing of data submissions and approximate sizes.
4. development of data documentation (metadata).
5. data disposition (path to archive center).
6. data affiliations, including both institutions and individual persons whose names will be associated with the data set in some way, *e.g.*, where did it come from, where does it go, *etc.*

These items are outlined for each of the PacIOOS platforms in the following sections.

#### A. Near-shore sensor (NSS)

1. List of parameters: ocean temperature, conductivity, salinity (computed), turbidity, florescence, and pressure (see Section V.C.4)
2. Processing steps: Data are sent from instrument via modem, converted to ASCII and entered into shore-side DataTurbine database at pre-set intervals (*e.g.*, 5 minute). Daily aggregated, netCDF files are made and served via THREDDS (see Section V.C.1).
3. Timing of submission: Annually submit collection of daily netCDF files for each sensor location; each location would be a separate “data set” under the aggregate “PacIOOS Near-shore sensors”.
4. Data documentation: The submitted netCDF files contain information on sensor location, instrumentation, points of contact, *etc.* (see Section V.C.3). Other documentation provided on the PacIOOS web site (<http://pacioos.org>).
5. Data disposition: Through the SIF development process, the exact location and mechanism of transfer of data between PacIOOS and NCEI will be determined. We understand that common practice for NCEI is to identify an FTP site, netCDF file format, SHA-256 checksum creation, and manifest file creation. We are prepared to implement this or similar procedures for each archived data stream and look to NCEI to help clarify the requirements (for file sizes see Section IV).
6. Data affiliations:
  - *Submitting institution*: PacIOOS
  - *Data point of contact*: Dr. Margaret McManus (mamc@hawaii.edu)
  - *Instrument technical point of contact*: Mr. Gordon Walker (gwalker@hawaii.edu)
  - *Data technical point of contact*: Dr. James Potemra (jimp@hawaii.edu)
7. Schedule: The SIF and initial transfer of the data of all NSS data collected through 2020 can happen as early as the last quarter of 2020 (our target is November), but no later than the first quarter of 2021. Subsequent transfers will be done annually in the first quarter of every year and will include data for the prior year.

## B. Water quality buoys (WQB)

PacIOOS supports a number of different Water quality buoys. At present there are two main principle investigators, Eric DeCarlo (Oahu buoys) and Jason Adolf (Hawaii Island buoys). Dr. DeCarlo has a separate arrangement with NOAA/PMEL to archive data, and PacIOOS no longer actively supports his group in terms of data services (although PacIOOS continues to serve the WQB data when provided). Thus, the archiving described below is presently limited to the Hawaii Island buoys.

1. **List of parameters:** salinity, dissolved oxygen concentration, ocean temperature, turbidity, chlorophyll, pH (see Section V.E.4)
2. **Processing steps:** Data from the Hawaii Island buoy(s) are provided by YSI via ftp, converted to ASCII and entered into DataTurbine database at pre-set intervals (e.g., 15 minute). Daily aggregated, netCDF files made and served via THREDDS (see Section V.E.1).
3. **Timing of submission:** Annually submit collection of daily netCDF files for each WQB location; each location would be a separate “data set” under the aggregate “PacIOOS Water Quality Buoy”.
4. **Data documentation:** Submitted netCDF files contain information on buoy location, instrumentation, points of contact, etc. (see Section V.E.3). Other documentation provided on the PacIOOS web site (<http://pacioos.org>).
5. **Data disposition:** Through the SIF development process, the exact location and mechanism of transfer of data between PacIOOS and NCEI will be determined. We understand that common practice for NCEI is to identify an FTP site, netCDF file format, SHA-256 checksum creation, and manifest file creation. We are prepared to implement this or similar procedures for each archived data stream and look to NCEI to help clarify the requirements (for file sizes see Section IV).
6. **Data affiliations:**
  - *Submitting institution:* PacIOOS
  - *Data point of contact:* Dr. Jason Adolf (jadolf@hawaii.edu)
  - *Instrument technical point of contact:* Dr. Jason Adolf (jadolf@hawaii.edu)
  - *Data technical point of contact:* Dr. James Potemra (jimp@hawaii.edu)
7. **Schedule:** The SIF and initial transfer of the data of all WQB data collected through already started. The older WQB data will be reprocessed and placed on the WAF for archive in late 2020. Subsequent transfers will be done annually in the first quarter of every year and will include data for the prior year

## C. High-frequency radars (HFR)

1. **List of parameters:** ocean surface current (see Section V.A.4)
2. **Processing steps:** Matlab binary files submitted directly to disk and converted to netCDF and ASCII each report cycle (e.g., 15 minute). Individual ASCII files pulled by the National DAC (Scripps/UCSD).
3. **Timing of submission:** TBD by DAC
4. **Data documentation:** Documentation provided on the PacIOOS web site (<http://pacioos.org>).

5. Data disposition: TBD by DAC.
6. Data affiliations:
  - *Submitting institution*: PacIOOS
  - *Data point of contact*: Dr. Pierre Flament
  - *Instrument technical point of contact*: Dr. Pierre Flament
  - *Data technical point of contact*: DAC
7. Schedule: TBD by DAC.

#### D. Gliders

PacIOOS currently has two gliders with only one operating at any one time. The main use of the glider data, aside from research activities, is to provide subsurface information for the numerical models. There have been limited missions thus far, order ten, and only the aggregate missions are served as historical data on the PacIOOS TDS.

1. List of parameters: depth, latitude, longitude, ocean temperature, salinity (computed), density, and pressure (see Section V.D.4)
2. Processing steps: PacIOOS glider data are reported in netCDF directly to NFS-mounted disks by the glider pilots. There is no further conversion done, but this format is very similar to the approved Glider DAC format. At present the glider data are only “released” to the PacIOOS modeling group, and there have not been many missions to date. Once missions are complete, the data are served via THREDDS (see Section V.D.1).
3. Timing of submission: TBD by DAC. PacIOOS will submit data to DAC.
4. Data documentation: Data documentation provided on the PacIOOS web site (<http://pacioos.org>).
5. Data disposition: TBD by DAC.
6. Data affiliations:
  - *Submitting institution*: PacIOOS
  - *Data point of contact*: Dr. Glen Carter
  - *Instrument technical point of contact*: Dr. Glen Carter
  - *Data technical point of contact*: Dr. James Potemra
7. Schedule: TBD by DAC.