

NODC Submission Information Form

(V1.3, Revised 01/2009)

FORM APPROVAL PENDING

Following the definitions and principles of the *Open Archival Information System (OAIS) Reference Model (ISO 14721:2003)*, this form documents the mutual understanding between a *Producer*, defined as a person or organization who provides information to be preserved, and an *Archive*, defined as the organization that intends to preserve information for access and use over the long term. It should accompany all data submissions to the National Oceanographic Data Center (NODC) and be completed to the extent possible.

The information contained on this form may be used to:

1. Populate NODC's Accession Tracking Data Base and product-specific databases
2. Create metadata records conforming to the Content Standard for Digital Geospatial Metadata (CSDGM), Vers. 2 (FGDC-STD-001-1998) and/or ISO 19115
3. Generate a formal archive appraisal package, for submissions requiring management level approval
4. Develop a list of *Producer* requirements requested of the *Archive*

The information contained on this form is true and correct to the best understanding of the *Producer* and *Archive* at the time of its submission. In the future, this information may be amended, updated, or revised as necessary and some submissions may require management level approvals before archival services can be provided.

Instructions:

This form is divided into six sections. Section 1 provides only the most basic Tracking Information and is the only section absolutely required at the time of submission. **However, within 1 month all submissions must also include information for Section 2**, which provides basic Data Discovery and Usage Information, and all submissions should strive to provide information through Sections 3 and 4, which provide more comprehensive and detailed information on the data set and its requirements for long term stewardship. Section 5 is required for submissions to the *Archive* that are expected to be periodic or routine in nature, and supports automation of archival services. Section 6 is optional and is only required for submissions that are expected to need management level approval and a formal archive appraisal package. The *Archive* Point of Contact will provide guidance as needed on all of these questions and will work with the *Producer* to ensure both parties reach a mutual understanding.

When complete, please email the signed form (see the last page of the document) to the *Archive* Point of Contact with a copy to NODC.DataOfficer@noaa.gov. Closing the email with "Signed," followed by your name is an acceptable form of signature.

Section 1 – Basic Tracking Information

NODC Submission Information Form

All elements in this section are REQUIRED, and will enable the *Archive* to establish a unique and durable tracking number known as an NODC Accession Number for the submission. It also clearly establishes whether the *Archive* is able to freely redistribute the data, and if not, what the restrictions are. When appropriate, the *Producer* will be provided the Accession Number in a confirmation receipt, along with the web address where the *Producer* can access the data set. When Section 1 is complete, the submission is considered “Initialized”.

1. Date of submission of this form (or its update):

2015-11-30 Mathew Biddle

2. Describe the scope of this data submission information.

PI Buoy Data from GLOS funded buoys

GLOS is one of 11 Regional Associations established nationwide through the NOAA Integrated Ocean Observing System (IOOS). IOOS coordinates the multi-agency, cooperative effort to routinely collect realtime data and manage historical information based on a continuously operating network of buoys, ships, satellites, underwater vehicles, and other platforms. These data are needed for many purposes which include rapid detection and prediction of changes in our nation’s ocean and coastal waters.

GLOS will be starting the automation process with a group of non-federal in-situ data sets. <http://glos.us/>

3. What is the Data Set title? A useful title includes a listing of two or three of the observed variables, the name of one or two of the platforms used to collect data or the project responsible for the data collection activity, the location, and the range of observation dates. For example, “Temperature, salinity, and nutrient data from bottle casts from the *Akademic Korolev, Alpha Helix, Polar Star, and Surveyor* in the Bering and East Siberian Seas from 1987-1999.”

Oceanographic and surface meteorological data collected from various stations by multiple institutions and assembled by Great Lakes Observing System (GLOS) near the Great Lakes.

4. Primary Point of Contact for *Producer* – please provide name, organization, position, address, telephone, fax, and e-mail address.

name: Tad Slawecki (administrative contact)
organization: LimnoTech (GLOS Contractor)
position: Senior Engineer
address: 501 Avis Drive, Ann Arbor, MI 48108
telephone: 734-332-1200
fax: 734-332-1212
e-mail: tslawecki@limno.com

name: Guan Wang (technical contact)
organization: Great Lakes Commission (GLOS contractor)

NODC Submission Information Form

position: *GIS Programmer/Analyst*
address: *2805 S. Industrial Hwy Suite 100, Ann Arbor, MI 48104-6791*
telephone: *734-971-9135*
fax: *734-971-9150*
e-mail: gwang@glc.org

5. Primary Point of Contact for *Archive* – please provide name, organization, position, address, telephone, fax, and e-mail address.

Mathew Biddle, Oceanographer
NOAA/NODC UMD/ESSIC/CICS E/OC1
1315 East-West Hwy
Silver Spring, MD 20910-3282
Phone: (301) 713-3272 X163
Email: Mathew.Biddle@noaa.gov

6. Can NODC freely and openly redistribute this dataset? If no, list the *Producer's* constraints of the Data Set in the *Archive* for Users in terms of:
- a. User access to the Data Set
 - b. Uses of the Data Set by Users

Yes

Section 2 – Basic Data Discovery and Usage Information

All elements in this section are REQUIRED if applicable and allow the *Archive* to enable users to find, access, and use the data described by this submission. It is strongly recommended that this information be provided at the time of the submission though it is acceptable if some elements are provided within **one month** of Initialization. When Section 2 is complete, the submission is considered “Active”.

1. What is the purpose for collecting this Data Set?

To collect realtime data and manage historical information based on a continuously operating network of buoys.

2. Provide a general descriptive abstract about the Data Set.

*PI Buoy Data from GLOS funded buoys
These are netCDF format in-situ buoy data collected in conjunction with GLOS. GLOS is one of 11 Regional Associations established nationwide through the NOAA Integrated Ocean Observing System (IOOS). IOOS coordinates the multi-agency, cooperative effort to routinely collect real-time data and manage historical information based on a continuously operating network of buoys, ships, satellites, underwater vehicles, and other platforms. These data are needed for many purposes which include rapid detection and prediction of changes in our nation's ocean and coastal waters.*

3. What is the time period covered by the Data Set?

We will be starting with the current data stream then move onto the historical once this process has been established.

4. What is the geospatial coverage of the Data Set (Easternmost longitude, westernmost longitude, northernmost latitude, southernmost latitude)? Note western longitudes and southern latitudes are negative, and use decimal degrees if possible.

Great Lakes

5. List the measured variables or parameters in the Data Set (e.g., Temperature, Salinity, etc.)

air_temperature, height, latitude, longitude, sea_surface_water_temperature, sea_surface_wave_significant_height, sea_surface_wind_wave_period, sea_water_temperature, time, wind_from_direction, wind_speed, wind_speed_of_gust

6. List the platform(s) from which the Data Set is derived.

NODC Submission Information Form

Buoys

7. List the instrument(s) used to derive the Data Set.

barometers, Temperature Sensors, Humidity Sensor, Aerovanes, bouy - moored buoy

8. List the observation types in the Data Set (e.g., Biological Data, Physical Data, etc.).

Physical data

9. List the mission/project name(s) to which the Data Set contributes.

IOOS, Integrated Ocean Observing System Data Assembly Centers Data Stewardship Program

10. Give the expected size(s) in bytes and number of files in the submission.

24GB

11. Give the file format and format version (e.g., netCDF-3, HDF-5, ASCII CSV, etc.).

netCDF-3 classic

12. Does this Data Set conform to any file-level data content or metadata content standards? (e.g., COARDS/CF, HDF-EOS, WOCE, GHRSSST)

netCDF version-3 classic CF-compliant and following ACDD

13. Please describe the file contents. Include enough information to make these data understandable to future users. For example, a table containing as applicable: parameter definition, data type, byte size/length, scale factor, offset, precision, and units. This information is especially important for ASCII and other formats which are not self-describing like netCDF and HDF. If this information is already contained in a file or file headers included in this submission, please indicate the file name.

In netCDF

14. Give the file-naming convention for the file(s) to be submitted, with the range/domain of each field value in the filename.

NODC Submission Information Form

Package Naming Structure

glos_buoyID_YYYY_MM_DD.tar.gz

<u>Term</u>	<u>Definition</u>
glos	Will always be listed as 'glos'
buoyID	Identifier for the specific buoy
YYYY_MM_DD	Date the package was generated.

File Naming Structure

buoyID_dataType.nc

<u>Term</u>	<u>Definition</u>
buoyID	Identifier for the specific buoy
dataType	Type of data that is assembled in the data file. (e.g. air_temperature)

15. Please provide a list of existing reports, publications, user guides, web sites, or other supporting documentation relevant to the Data Set.

<http://glos.us/>

16. What metadata exists for this Data Set? Is it in a standard format/can it be automatically translated into a standard format? Describe the granularity of this metadata (For example, is it collection level metadata? If not, to what file or grouping of files does it apply?)

Inside the netCDF files to the data type level

17. If applicable, describe the temporal resolution of the primary parameter(s) in the Data Set.

~10 mins at highest resolution

18. If applicable, describe the horizontal resolution of the primary parameter(s) in the Data Set.

~1/100th of a decimal degree, 0.01 Degree OR 1" in DD° MM' SS" format

19. If applicable, describe the vertical resolution of the primary parameter(s) in the Data Set.

NODC Submission Information Form

~1/100th meter, .01 meter

20. If applicable, describe the projection grid or coordinate system used in the Data Set.

```
crs:inverse_flattening = 298.257223563 ;  
crs:epsg_code = "EPSG:4326" ;  
crs:long_name = "http://www.opengis.net/def/crs/EPSSG/0/4326" ;  
crs:grid_mapping_name = "latitude_longitude" ;
```

21. If the Technical Contact for the *Producer* is different from the Primary Contact for the *Producer* (1.4), please provide name, organization, position, address, phone, fax, and email.

Same as above.

22. If the Metadata Contact for the *Producer* is different from the Primary Contact for the *Producer* (1.4), please provide name, organization, position, address, phone, fax, and email.

Same as above.

23. If the Technical Contact for the *Archive* is different from the Primary Contact for the *Archive* (1.5), please provide name, organization, position, address, phone, fax, and email.

See above

24. If the Metadata Contact for the *Archive* is different from the Primary Contact for the *Archive* (1.5), please provide name, organization, position, address, phone, fax, and email.

See above

Section 3 – Detailed Data Processing and Quality Information

All elements in this section are STRONGLY RECOMMENDED and allow the *Archive* to enable more complete and thorough understanding of the data over the long term.

1. What is the overall completeness and quality of the Data Set?

Submitted data have been subjected to automated QC, including range and missing data checks.

2. Describe the data processing level of the Data Set. For example, is the Data Set unprocessed or minimally processed, quality controlled or calibrated, etc.? For satellite data, is it Level 0, 1, 2, 3, or 4?

NODC Submission Information Form

Automated QC checks for data availability and data range have been performed on submitted data.

3. Summarize the science algorithms(s) used to derive the Data Set.

None.

4. Describe the steps taken to process the Data Set, including for each step the methodology, source data, and time/frequency, and listing any input data sets used to derive the Data Set.

None.

5. Describe the Data Set's dependency on other data (e.g. ancillary files), processing systems, software, or entities that are not to be submitted to the Archive.

GLOS uses a hybrid Lua and C software to generate the netCDF files. Lua is used from the metadata management aspect and C is used for generating the netCDF files with the metadata processed by Lua.

6. Detail any measures taken by the Producer to assess the quality of the Data Set, including data comparisons, and an assessment of the attribute accuracy. Give information about omissions, selection criteria, and other rules used to derive the Data Set.

Data uploaded to GLOS from sensors are subjected to automated availability and range checks. If data are not available, a missing value placeholder is inserted in the data record. If a data value is outside of the acceptable range, the data value is replaced with the appropriate extrema of the data range.

7. List any quality assessment parameters included in the Data Set. For example, this may be an explanation of quality flags and their range/domain of values.

None at this point. GLOS anticipates implementation of QARTOD (or similar QC for non-QARTOD parameters) by 12/31/2017, at which point QC flags are expected to be included in archive submittals.

Section 4 – Data Stewardship Information

All elements in this section are STRONGLY RECOMMENDED and enable the *Archive* to provide more comprehensive *data stewardship* over the long term. Data stewardship requires a more extensive set of functions than traditional long-term preservation of data and information, and includes activities such as monitoring the needs of user communities, compliance testing, quality assurance, and use of this Data Set in larger integrated product databases. Importantly, this section provides the *Producer* with an opportunity to request specific services from the *Archive*. This document does not imply that all of these services will be provided, but typically the *Archive* will work to meet them on a best-effort basis.

1. Please describe any quality control or quality assurance procedures the *Archive* should perform on this Data Set when it is submitted to the *Archive*. How will the *Producer* provide updates to the *Archive* when changes occur in the Data Set, transmission mechanism, format, content, etc.? How often might such changes be expected to occur?

The archive will verify the data has been downloaded correctly with md5 checksums. GLOS will notify NODC of any changes to the automation through e-mail correspondence with Mathew.Biddle@noaa.gov. We don't expect changes to occur. Any changes must be clarified with NODC.

2. Does the *Producer* request reports on the *Archive*'s dissemination of the Data Set? If so, what statistics should be included? (Please note federal regulations strictly limit the amount and kind of information that can be recorded by federal agencies.)

Yes, data usage statistics. As well as when NODC begins the ingest processing.

3. Does the *Producer* request standards compliance testing on the Data Set? For example, should the *Archive* verify data files are meeting netCDF Climate and Forecast (CF) conventions, or should metadata records be checked for adherence to the FGDC content standard? Will the *Producer* perform standards compliance testing prior to submission to the *Archive*?

Yes, NODC will check for CF compliance with standard_names. As well as general netCDF formatting.

4. Suggest action(s) for the *Archive* in the case of an error in transmission (e.g. missing data, duplicate data, incorrect file name or size, failure of compliance checks).

Contact GLOS through e-mail with a description of the problem that was encountered.

5. Please list any known NODC product databases (e.g. World Ocean Data Base) that this Data Set should become a part of.

NODC Submission Information Form

None known of now. Possibly Marine Environmental Buoy Database.

6. Please identify one or more Representative Users of the *Designated Community*. The Designated Community is defined in the OAIS Reference Model as the group of potential users who should be able to understand a Data Set over the long term. The *Archive* works specifically to preserve the data and information for this Designated Community.
- Describe this user community and their requirements
 - Provide Contact Information for a representative of this community - please provide name, organization, position, address, telephone, fax, and e-mail address

?

7. List security requirements for dissemination of the Data Set from the *Archive* to the users.

None.

8. Once the Data Set is transferred to the *Archive*, how long should it take for it to become searchable? How long should it take to become accessible online?

Immediately.

9. Describe any preferred search criteria to be enabled for this Data Set in the *Archive* (e.g., search by time, search by geographic bounding box on a Polar Stereographic map, etc.)

Geoportal search capabilities.

10. Describe any the preferred access mechanisms to be enabled for this Data Set in the *Archive* (e.g., OPeNDAP, Web Coverage Services, FTP, etc.)

FTP, OPeNDAP, THREDDS WCS, WMS, HTTP, any others applicable to netCDF data files.

NODC Submission Information Form

Section 5 – Logistics Information for Routine Transfers to the *Archive*

All elements in this section are REQUIRED for Data Sets that are expected to be routine, automated transmissions to the *Archive* from the *Producer*. This information is required for the *Archive* to establish and maintain the automated ingest and archive procedures. Questions 1 through 5 in Section 4 above are also required for automatic submissions.

1. Provide the mechanisms used to transfer digital data to the *Archive*. For routine, repeated submissions include the server, location, and protocol used.

NODC will pull data from the GLOS FTP site at glos.us. The data will be pulled monthly and checked against the accompanying manifest file.

2. List any relevant Interface Control Document, Memorandum of Understanding, or other technical documents outlining how data will be transferred from *Producer* to *Archive*.

None, for now...

3. Describe the submission schedule in terms of starting/ending times and submission frequency for each submission session.

The archive will check the FTP site at glos.us for any new files/updates to files once per month. The ingest and archival of the checked data will happen subsequent to the checking of the FTP site.

4. Give the volume of each submission session and the total anticipated volume per day or month in bytes.

Anywhere from 60KB to 900KB unpacked and 4KB to 400KB packed

5. List the steps in the transfer process from *Producer* to *Archive*.

NODC will check GLOS' FTP server for new files meeting the filename convention each month. If a new file appears, it will be copied to a working directory on NODC's server and prepared for ingest. If a file already exists in the archive but the checksums have changed, the previous file will be updated with the most recent submission. Each package will be its own archival package, the packages are based on buoy identifiers, so each package will be its own archival package.

6. List the *Producer*'s preference for basic file validation routines (e.g. checksums, CRC32, MD5 or other).

NODC Submission Information Form

MD5 checksum

7. Does the *Producer* request a periodic record of receipt from the *Archive* for purposes of tracking the submitted data?

Yes, GLOS would like updates once we copy the data over and when the data have been archive-published.

8. List any security requirements needed during submission from the *Producer* to the *Archive*.

None

9. Is the content of each submission session considered by the *Producer* to be a continuation or new version of a previous submission, or is the content of each submission session considered by the *Producer* to be an independent or stand-alone collection of data?

For the first time the automation runs, it will generate new accessions for each Bouy ID. If the submission is for the same Buoy ID, then an update to a previous version of that accession should be conducted.

Section 6 – Archive Appraisal and Justification Information

Only in cases where a formal archive appraisal package is required by the *Archive* in order to gain management approval to provide archival services for this Data Set are these elements required. For these cases, also ensure the following questions have been answered: all of Section 2, and Section 3 questions 1 and 2. The Point of Contact for the *Archive* will provide additional guidance with this section.

1. What are the cost considerations for long-term maintenance of the Data Set? Are resources available for archiving and providing access to these records?

This data falls within the scope of NODC and does not incur any additional costs. Resources are currently available for archiving and providing access to these records.

2. Has this Data Set ever physically resided at a scientific data center or center of data where stewardship was provided? Where does it reside now? What scientific expertise would best provide stewardship for this Data Set?

No, this data has not physically resided at a scientific data center.

3. Where does this Data Set fit within NOAA’s mission?

Yes.

4. What is the value (scientific, public, government) of this Data Set in terms of current and anticipated future benefits?

?

5. Does the Data Set have legal mandates which require its archive at NOAA? Are there existing NARA disposition schedules that pertain to these records? If yes, please describe.

No.

6. Is the Data Set unique? If not, where else does it exist?

Yes.

7. Is the Data Set related to other records in a NOAA *Archive* (i.e. an extension, a new version, improved quality, etc.)? If yes, to what degree does this Data Set add value to other data sets held by NOAA or others?

NODC Submission Information Form

No.

8. Has the Data Set undergone user evaluation and/or scientific peer review, been used extensively in publications, and/or subjected to other appraisal processes? If yes, please describe.

No.

9. What is the current storage media for the Data Set? If in electronic format, does it still exist on other media (e.g. paper, film)? If yes, is it required to maintain copies on other media?

Only electronic format, specifically netCDF.

10. Does appropriate hardware and software technology exist to enable usability of the Data Set? If yes, please describe.

Yes. Any netCDF compliant software (e.g. Matlab).

11. Does the Data Set have intrinsic value? Intrinsic value implies that an object containing data has value beyond the data content in the object. For example, the original deck logs from the HMS *Beagle* have intrinsic value, but the digitized observations from those logs do not because the digitized files are easily copied viewed, and/or redistributed.

No.

The signatures below indicate the belief that the information contained on this form is true and correct to the best understanding of the *Producer* and *Archive*. These signatures also acknowledge that in the future, this information may be amended, updated, or revised as necessary and that some submissions may require management level approvals before archival services can be provided.

Theodore A.D. Slawewski

Point of Contact for the *Producer*
Printed Name and Date:

Theodore A.D. Slawewski
August 24, 2015

Mathew M. Biddle

Point of Contact for the *Archive*
Printed Name and Date:

Mathew M. Biddle
August 24, 2015

NODC Submission Information Form

Updated: 2015-11-30 Theodore A.D. Slawewski

Updated: 2015-11-30 Mathew M. Biddle