

NODC Submission Information Form

(V1.3, Revised 01/2009)

FORM APPROVAL PENDING

IOOS National Glider Data Assembly Center

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

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):

Updated on 2016-02-02, Datastream went operational 2015-12-01

2. Describe the scope of this data submission information.

Glider data submitted to NCEI by the IOOS National Glider Data Assembly Center (NGDAC)

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.”

Physical trajectory profile data from glider “platform id” deployed by “institution” in the “sea_name” from “time_coverage_start” to “time_coverage_end” (NCEI Accession XXXXXXXX)

The “platform id”, “institution”, “sea_name”, “time_coverage_start”, “time_coverage” for each Archival Information Package (AIP) are extracted or calculated from the data in the netCDF files.

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

John Kerfoot
Rutgers University
71 Dudley Road - Office 111B
New Brunswick, NJ 08901
office: (848) 932-3344
cell: (732) 687-9877
kerfoot@marine.rutgers.edu

5. Primary Point of Contact for *Archive* – please provide name, organization, position, address,

telephone, fax, and e-mail address.

Tom Ryan
NESDIS/NODC/DBMD
Oceanographer
Building: SSMC3 Room 4715
Silver Spring, MD 20910
Phone: 301-713-4916
Fax: 301-713-3302
Email: thomas.ryan@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?

The purpose of collecting this data set is to respond to the need for subsurface oceanographic observations in the US coastal and other areas.

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

The paragraph below will be added as a second paragraph to AIPs. The first paragraph will be extracted directly from the “summary =” field of the deployment netCDF file.

The National Centers for Environmental Information (NCEI) received the data in this archival package from the Integrated Ocean Observing System's National Glider Data Assembly Center ([IOOS](#) NGDAC). The [IOOS](#) NGDAC received the data in one or more netCDF files comprising an entire glider deployment. The data are measurements of physical oceanographic properties such as temperature, salinity, conductivity, and density. The [IOOS](#) NGDAC checked the files for compliance to their netCDF file convention, aggregated the files into a single netCDF file, and then submitted the file to NCEI for long-term preservation.

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

2013-01-22 through present

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.

Easternmost longitude = 180
Westernmost longitude = -180
Northernmost Latitude = 90
Southernmost Latitude = -90

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

As of 2016-02-02 variables as below. Additional variables will be added over time.

sea_water_electrical_conductivity, depth, sea_water_pressure, sea_water_salinity,
sea_water_temperature, sea_water_density, eastward_sea_water_velocity,
northward_sea_water_velocity, longitude, longitude, time

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

As of 2016-02-02 glider platforms as below. Additional platforms will be added over time.

bass, blue, ru01, ru04, ru05, ru22, ru23, ru24, ru26d, ru28, SG609, SG610, sp001, sp010,
sp011, sp018, sp020, sp025, sp028, sp030, sp031, sp035, sp039, sp040, sp042, sp043, sp047,
sp048, sp050, sp051, sp052, sp063, ud_134, unit_308, unit_540

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

As of 2016-02-02 instruments as below. Additional instruments will be added over time.

Sea-Bird 41CP
Seabird GPCTD
unpumped Seabird SBE41
Seabird SBE 41CP

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

As of 2016-02-02 observation types as below. Additional observation types will be added over time.

Physical

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

As of 2016-02-02 missions/projects as below. Additional missions/projects will be added over time.

LTER - Long-Term Ecological Research Program
CONVERGE - Impacts of local oceanographic processes on Adélie penguin foraging ecology
COLLABORATIVE RESEARCH: The Propagating Response of the Inner Shelf to Wind
Relaxations in a Coastal Upwelling System
MARACOOS - Mid-Atlantic Regional Association Coastal Ocean Observing System
Sustained Ocean Observations for Improving Atlantic Tropical Cyclone Intensity and
Hurricane Seasonal Forecast

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

One file per submission, ~40 MB

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

CF compliant netCDF-4

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

CF compliant netCDF-4

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.

Files submitted for archive are in CF compliant netCDF-4 format which is self describing

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

<ggggg>_YYYYMMDDTHHMM.ncCF.ncV.nc
where <ggggg> = glider id; YYYYMMDD = numeric year, month, day of deployment;
T = time; HHMM = numeric hours and minutes of deployment;
ncCF = CF compliant netCDF; ncV = nc indicates netCDF and V is the netCDF
version number; nc = netCDF file extension.

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

IOOS NGDAC wiki - <https://github.com/ioos/ioosngdac/wiki>

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?)

An extensive set of metadata is included as global and variable attributes in the netCDF files.

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

varies

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

varies

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

varies

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

WGS84

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 primary contact

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 primary contact

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.

John Relph
Organization: NCEI
Position: IT Specialist
Address: SSMC 3, Room 4842
1315 East-West Highway
Silver Spring, MD 20910
Phone: 301-713-4914
Email: John.Relph@noaa.gov

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.

Same as primary contact

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?

Complete and of high quality

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?

Varies between Level 1 and 2

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

Glider profile lag correction, practical salinity (1978), sea water density

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.

Glider collects data from attached instruments and transfers them shoreside (frequency varies). Shoreside collection data logger performs calibration corrections and basic quality control and creates log files (frequency varies).
Data provider translates log files into GliderDAC compliance netCDF files (frequency varies).
Glider DAC aggregates individual files and publishes them (hourly).

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.

Dependent on ERDDAP for aggregation of files.

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 logger performs basic quality controls (range checks).
Additional quality control performed by the data originator varies.
ERDDAP and the IOOS Compliance Checker are used to assess the attribute accuracy.

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.

The Glider DAC follows IOOS adopted QARTOD standards for quality control.

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?

Check for completeness of transmission and notify provider if problems occur.

- a. Check for package completeness (manifest check with versions)

2. 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?

Changes should be infrequent; however they will occur over time. Change in netDCF manifest checksum will signal update to a submission.

3. 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.)

Not required

4. 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*?

Not required

5. 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).

A email will be sent to ioos.glider.data@noaa.gov outlining the error or problem.

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

World Ocean Data Base

7. 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

- Glider operators and modelers that assimilate glider data
- John Kerfoot
Rutgers University
71 Dudley Road - Office 111B
New Brunswick, NJ 08901
office: (848) 932-3344
cell: (732) 687-9877
kerfoot@marine.rutgers.edu

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

None required

9. 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

10. 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.)

No preference

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

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.

NCEI will retrieve a copy of each file via FTP protocol.
The SIP is retrieved via sftp by signing into from ammonite at:
sftp NCEI@data.ioos.us and cding to the archive/ directory.
Note: Public and private keys were generated by Relph and Relph has the password for log in.

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

This document will stand on its own.

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

The sftp site is accessed daily and all new Submission Information Packages (SIPs) are retrieved.

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

Each submission is ~40 MB and the anticipated volume per day will be variable.

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

The NCEI archive(s) servers will routinely query the FTP account for new md5 files. When a md5 file is found the corresponding netCDF files will be copied to the archive(s) and verified using the md5 checksum.

Verification is accomplished by comparing the md5 checksum file hex string to the corresponding one contained in the md5 file.

NCEI will notify the NGDAC only if negative issues arise.

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

MD5 or other).

MD5

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

A publication email will be sent to ioos.glider.data@noaa.gov upon publication of all AIPs.

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

No

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?

stand-alone

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?

No budget available.

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

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

This data set supports NOAA's mission of Science, Service and Stewardship by providing coastal observations to the public. It directly supports the NOAA Goal of Resilient Coastal Communities and Economies.

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

It is anticipated these these data will be used extensively in scientific studies by many categories of users, particularly to study climate variability and change. As the records become longer over time their value will be even greater.

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?

Data are currently being served by the IOOS National Glider Data Assembly Center (NGDAC) via an ERDAP server - <http://data.ioos.us/gliders/erddap/info/index.html?page=1&itemsPerPage=1000>

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?

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.

It is anticipated that these data will be frequently cited in scientific publications and will be evaluated by many scientific investigators.

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?

Electronic

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

There are many applications available that handle netCDF.

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.

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

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