AOSN Data Policy

Autonomous Ocean Sampling Network

22 July 2004

Philosophy and Motivation

The success of the AOSN program depends on the collaboration of a network of institutions geographically spread across the globe. Our objectives require quantitative analysis of interdisciplinary data sets and therefore data must be exchanged between researchers. To extract the full scientific value, data must be made available to the scientific community on a timely basis.

As stated in the *AOSN II System Goals and Performance Metrics*, achieving Operational Objective A2 (data flow), requires a data sharing strategy:

> "A centralized system will collect data/products, make them available through the web, and will archive all versions of data. Originators (observing elements or modeling groups) will be primarily responsible for quality control of their own data/products. To assist in improving QC procedures, users will inform the control room of expected data/product-quality issues. The control room will solicit information and/or corrective action from data originators."

Pursuing such a strategy means that as soon as data might be useful to other researchers, the data should be released, along with documentation which can be used by the other researchers to judge data quality and potential usefulness. The value of such a policy is well stated in the U.S. Globec Data Policy:

> "…The intellectual investment and time committed to the collection of a data set entitles the investigator to the fundamental benefits of the data set. Therefore, publication of descriptive or interpretive results derived immediately and directly from the data is the privilege and responsibility of the investigators who collect the data. The purpose of a data archive is to facilitate collaboration between scientists, the combination of multiple data sets for interdisciplinary and comparative studies, and the development and testing of new theories."

However, such a data sharing system depends on prudent use of the data by other investigators. The definition of appropriate use depends on the nature of the data, and the relationship of the data generator to that data. In some cases, access to raw numerical data may only be granted to individuals willing to register themselves as data users, and who acknowledge the data use policy. Also, publication of the data may require a citation, contact with the author, or even permission from the author. Hence, important sets of metadata in the AOSN data repository are the actions required to access the data and/or to publish results from using the data.

To facilitate communication, administrators of AOSN data repositories will maintain a list of all data access and will notify those who access the data of our commitment to the principle that data is the intellectual property of the collecting scientists.

Actions Required to Publish Data

Users of AOSN data who wish to publish their results may have to perform up to three actions, depending on the data set:

1) Provide citation to the data:

   a) No citation required – some data sets are "institutionalized" data that are beyond the control of the data originator. An example of such data includes National Weather Service data. While a reference to such data is considered optional, a reference is encouraged and the data repository will contain a reference for data users to adopt.

   b) Citation required – in this case, the data creator may have a vested interest in the data, and would like to receive acknowledgement for its use through a reference. Alternatively, revisions to the data occur from time to time, and the appropriate data revision must be referenced. The data repository will contain a reference for data users to adopt.

2) Obtain permission from the data creator:

   a) No permission required.

   b) Permission required -- the uniqueness, complexity, or revision status of the data is such that further discussion is required before the data can be published. Alternately, institutional data policies may dictate permission and other actions as a prerequisite to publishing. The data repository will contain the a contact name for each data set that requires permission prior to publication.

3) Expectation of co-authorship:

   a) No expectation of co-authorship.

   b) Co-authorship expected – scientists making use of such a data set should anticipate that the data collectors might require co-authorship of published results. The data repository will contain a name to contact regarding co-authorship.

Citations

References to data obtained from the AOSN program will depend on the form of the distributed data (on-line, CD-ROM, or DVD). The data set citation will be unique to each data set, and will be agreed to by the principal investigator or primary point of contact for the data set.

Citation for Downloaded Data (APA styles are shown here):

   [Data set citation] *Autonomous Ocean Sampling Network (AOSN) 2003 Field Experiment: AOSN 2003 Online Data*. Retrieved [retrieval date], from the AOSN project Web site: http://aosn.mbari.org

Citation for Data on CD-ROMs:

   [Data set citation] *Autonomous Ocean Sampling Network (AOSN) 2003 Field Experiment: CD-ROM Data*. Moss Landing, CA: Monterey Bay Aquarium Research

Institute. Retrieved from Autonomous Ocean Sampling Network (AOSN) 2003 Field Experiment (AOSN 2003 CD-ROM Data, CD-ROM, [release date] – [volume#])

Citation for Data on DVDs:

[Data set citation] *Autonomous Ocean Sampling Network (AOSN) 2003 Field Experiment: DVD Data*. Moss Landing, CA: Monterey Bay Aquarium Research Institute. Retrieved from Autonomous Ocean Sampling Network (AOSN) 2003 Field Experiment (AOSN 2003 DVD Data, DVD, [release date])

Roles and Responsibilities

<u>Primary investigators</u> are responsible for either uploading their quality-controlled data to the Polarbear server at MBARI, or placing it (in a COARDS-compliant format) on-line in a DODS (OpenDAP)-accessible database. In either case, investigators are also responsible for alerting AOSN staff at MBARI that the update has taken place.

<u>AOSN staff at MBARI</u> will make the data on Polarbear and data hosted on external DODS (OpenDAP) servers accessible via the AOSN Live Access Server. All data accessed through the AOSN Live Access Server will be presented with consistent variable names and unit formatting (using aliases as required), to foster the inter-comparison of different data sets.

The Live Access Server maintained at MBARI will require users to enter their e-mail other contact information to access the server. They will then receive a machine-generated password that will allow them to download all data as images but only Use Level 1 data in numerical formats. Users who wish to download Use Level 2 and 3 data in numerical formats will have to request an access upgrade, and will be asked to agree to acknowledge the data originator in their use of the data. Access upgrades will require the approval of a member of the AOSN staff.

CR-ROM and DVD versions of the data available through the AOSN Live Access Server and OpenDAP server will be periodically published by MBARI, to meet the needs of data users who may not be allowed to reference on-line data sources.

Ratification and Exemptions

The AOSN Executive Committee will ratify this data policy. Researchers requesting exemption from the data policy must obtain a consensus agreement on the exemption from the AOSN Executive Committee.