Interagency Ocean Acidification Data Management Plan: Draft One

June 23, 2012

INTRODUCTION

Ocean Acidification refers to the chemical changes happening in marine and estuarine waters as a result of rising CO_2 in the atmosphere. The pH of the world oceans is decreasing as they absorb approximately one third of all CO_2 emitted by humans activity (Sabine et al 2004). The decrease in pH is also leading to a decrease in saturation state of certain carbonate minerals important for shell formation in some marine organisms. As a result of the rising concern about ocean acidification, it is critically important that researchers around the world have easy access to diverse, relevant data ranging from observing data (from fixed moorings and dedicated cruises) to the results from experiments testing the impact of rising CO_2 on marine organisms to model output. It is also important that the general public have access to data synthesis products for general understanding of this phenomenon. Synthesis product development (such as near-real-time availability of data relevant to shellfish growers or coral reef managers that indicate the current and near-term trends in carbonate chemistry in a local region) will rely on access to data from multiple sources.

According to the recently released *Draft* Interagency Working Group Strategic Research Plan on Ocean Acidification: "The success of the National Ocean Acidification Enterprise will depend critically on effective data management and integration. Data must be shared and integrated across disciplinary boundaries, drawing marine biological data together with oceanographic data and providing intelligible information to social scientists, planners, educators, and the general public." The SRP was mandated by the Federal Ocean Acidification Research and Monitoring (FOARAM) Act of 2009. Section 12404 b(5) of the FOARAM Act requires the USG to "establish or designate an Ocean Acidification Information Exchange to make information on ocean acidification developed through or utilized by the interagency ocean acidification program accessible through electronic means, including information which would be useful to policymakers, researchers, and other stakeholders in mitigating or adapting to the impacts of ocean acidification."

Seeing a critical need for coordination and as lead for the Interagency Working Group on Ocean Acidification, the NOAA Ocean Acidification Program (established May 2011) has taken the lead to integrate ocean acidification-related data across the federal government, recognizing the paramount importance for data sharing both nationally and internationally. As a result, an OA Data Management Workshop was convened in March 2012 at the University of Washington under the auspices of the NANOOS/PNW regional association of IOOS. Representatives from across NOAA, from other ocean-related science federal agencies and from academia attended representing expertise in data management and scientific research experience. Through presentations of the state of data management and lengthy discussion in small groups a basic operational plan for working together moving forward was developed. *This document represents that plan, conceived in the workshop. We must consider it a living document with more questions posed than answered. The plan should be vetted through the CIMOAD. The issues presented in the appendices still need to be incorporated into the main body of the plan.*

Strategic Vision and Declaration

The ocean acidification science community is purposefully diverse, and the data being collected is equally heterogeneous, spanning experiments, sustained ocean monitoring, satellites, and models. The challenge of integrating these data sources for broad application requires a cooperative approach between scientists and data managers. The OA Data Integration Framework needs to both develop new and build on existing relationships between scientists and data managers. Associated roles and responsibilities of these partners are distinct while also complimentary. Given limited or shrinking financial resources, it is critical that these relationships be enhanced. To that end, the emerging ocean acidification community has developed a so-called Declaration of Interdependence, which articulates the goals and vision for ocean acidification data integration in the U.S with specific, actionable recommendations for interagency action.

<u>What an OA Integrated System might look like!</u>

- Living track back to the data and metadata—"Car Fax" style search/Kayak/Provenance.
- System that continually informs users, generators and institutions what is going on with a particular data set
- Management systems get ranked with how much they get used—impact factor
- Interactive site—tells users that there are other related data sets
- Well organized = searchable
  - Clearly and simply defined
  - User experience must be positive
  - Data delivery system that can respond to the user—keeping in mind that this is a multi-user platform
  - Maximum automation for users for easy upload—easy submission (simple and fast)
  - Standardized data formats

“**Declaration of Interdependence of Ocean Acidification Data Management Activities in the U.S.**”

Whereas Ocean Acidification (OA) is one of the most significant threats to the ocean ecosystem with strong implications for economic, cultural, and natural resources of the world;

Whereas our understanding of OA and our ability to: 1. inform decision makers of status, trends, and impacts, and 2. to research mitigation/adaptation strategies, requires access to data from observations, experiments, and model results spanning physical, chemical and biological research;

Whereas the various agencies, research programs and Principal Investigators that collect the data essential to understanding OA often pursue disparate, uncoordinated data management strategies that collectively impede effective use of this data for synthesis maps and other data products;

Whereas an easily accessible and sustainable data management framework is required that:
i) provides **unified access to OA data for humans and machines**; ii) **ensures data are version-controlled and citable** through **globally unique identifiers**; iii) documents and communicates **understood measures of data and metadata quality**; iv) is easy to use for submission, discovery, retrieval, and access to the data through a **small number of standardized programming interfaces**;

Whereas urgency requires that short-term actions be taken to improve data integration, while building towards higher levels of success, and noting that immediate value can be found in **the creation of a cross-agency data discovery catalog** of past and present OA-related data sets of a defined quality, including **lists of parameters, access to detailed documentation, and access to data via file transfer services and programming interfaces**;

Whereas this integration will also benefit other users of data for a diverse array of investigations;

Therefore, be it resolved that the 30 participants of an OA Data Management workshop in Seattle, WA on 13-15 March 2012 established themselves as the Consortium for the Integrated Management of Ocean Acidification Data (CIMOAD) and identified three necessary steps forward to achieve this vision:

1. The **endorsement of agency program directors and managers for collective use of machine-to-machine cataloging and data retrieval protocols (including THREDDS/OPeNDAP)** by each agency data center to provide synergistic, consolidated mechanisms for scientists to locate and acquire oceanographic data;

2. The commitment of the scientific community to **establish best practices for OA data collection and metadata production**, and the leadership to provide a means of gaining this consensus; and

3. The endorsement of agency program directors and managers to direct data managers to collaborate to develop the system articulated above and **contribute to a single national web portal to provide an access point and visualization products for OA**.

We, the undersigned, request your attention to this matter and commitment to bringing this vision to reality in the next five years for the benefit of our nation and contribution to the global understanding.

Implementation of the vision

The overarching activities, above, will not come to fruition without concrete plans that must include short-term opportunities for progress arising from existing projects, priorities and funding.

Each distinct scientific community (observing, experimental, modeling, and satellites) must self-identify (Fig. 1) and ensure a coordinated approach for development of content and formats of data and metadata and defined quality control procedures that are both human and machine-readable, with standardized units and variable names, and metrics to indicate completeness of metadata.

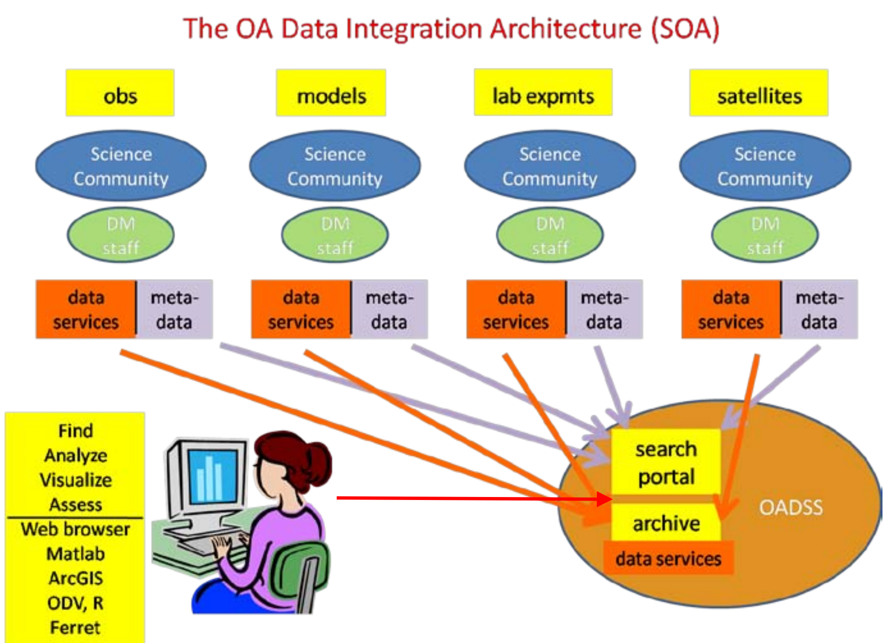


Fig. 1. Proposed operational structure of an integrated OA data management system. User can access data from a variety of data providers through one central access point.

Data management staff must work to bridge the scientific communities, to agree on data access services and a strategy for data citations, translate scientific metadata content into industry standards for optimal discovery, and make the data available, with clearly designated levels of quality control, using agreed upon web services.

Ultimately, the science and the data need to be coordinated and accessible to a broad array of users and applications. A focal point is needed to effectively maintain a search portal that provides discovery and access to preserved (archived) data that span the scientific communities that comprise ocean acidification science. An effective Ocean Acidification Data Stewardship

System (OADSS) will ensure that the OA data can be used with users' applications and provide assistance to direct users to the necessary data and products.

Characteristics of an OA Data Stewardship Framework as led by NODC for NOAA (and US)

- Adaptable framework
- Robust rich metadata and data history (PIs keep track of data evolution)
- Easy to integrate OA data with other relevant measurements (data products)
- Archival data granularity
- Unique ID (accession/DOI?) + data version control (coordination)
- Easy data discovery and access services (NetCDF templates)

Recommended near-term plans:

Step A: The endorsement of agency program directors and managers for collective use of machine-to-machine cataloging and data retrieval protocols (including THREDDS/OPeNDAP) by each agency data center to provide synergistic, consolidated mechanisms for scientists to locate and acquire oceanographic data.

Action Item A1: Data managers across the OA scientific community must agree upon the desired data cataloging and data retrieval protocols. These must be articulated as an appendix to this evolving data management plan.

Action Item A2: The requirements for cataloging and data retrieval must be communicated to the funding agencies, to be incorporated into funding opportunity announcements.

Step B: The commitment of the scientific community to establish best practices for OA data collection and metadata production, and the leadership to provide a means of gaining this consensus.

Action Item B1: With assistance from OADSS staff, each scientific community (PIs plus data managers in each of 1) observing, 2) experimental (both laboratory and in situ) and 3) modeling components) will identify a lead to coordinate and develop a plan for the development of defined data collection formats and metadata content and common quality control procedures and flags. It might be useful to establish a "data test bed" by focusing in on one community or subset of projects before tackling entire data flow.

Action Item B2: The National Oceanographic Data Center will establish OADSS and lead the development of data management guidelines and metadata content standards.

Action Item B3: NODC/OADS will work with the scientific community to define what comprises an ocean acidification data set and create a defined vocabulary for the parameters included.

Step C: The endorsement of agency program directors and managers to direct data managers to collaborate to develop the system articulated above and contribute to a single national web portal to provide an access point and visualization products for OA.

Action Item C1: Funding announcements will include guidance consistent with this and other community plans for coordinated data collection and documentation.

Action Item C2: NODC/OADSS will deploy the agreed upon web services and search portal to enable discovery of and access to all OA data. Visualization tools will be developed where possible.

Action Item C3: NOAA will develop a procedural directive to ensure a means for citation of OA and other oceanographic data which likely include an established DOI (Digital Object Identifier) procedure.

Step D: Get word out about data management next steps identified in this workshop. Look for near term meetings, such as the Oceans in High CO2 World symposium, for opportunities to broadcast this message. Also, it is important that OCB-OA subcommittee play a role in data management practices and dissemination.

Appendix 1

Current data management capabilities

An overview of current capabilities from various data management efforts which are or could be directed to ocean acidification data, as presented by workshop attendees, is presented below. Included is information on data management entities, data streams, examples of ocean acidification variables, Quality Assurance (Qa) and Quality Control (Qc) processes, data serving tools, and data archival strategies. Other capabilities, as identified, can be added to this document.

Table 1: Basic information on major data management efforts, funded by or managed by the federal government, which now or may include *ocean acidification* as a special emphasis

Program Name	**Federally funded Data Management Players**
NOAA NODC	NODC will serve as NOAA's Ocean Acidification data management focal point by providing dedicated online data discovery, access, and long-term archival for a diverse range of OA data. NODC is the designated federal permanent archive for chemical, physical, and biological oceanographic data
IOOS	A national-regional partnership working to: • Enhance our ability to collect, deliver, and use ocean information • Provide new tools and forecasts to improve safety, enhance the economy, and protect our environment Integrate data from a wide diversity of sources and providers Encourage and support strategic partnerships (thematic, technological, regional) Augment the OA monitoring network by encouraging/enabling IOOS RA platforms to be included within the OA monitoring strategy. Platforms of opportunity, reducing duplication.
BCO-DMO	Mandate: is to provide data management support throughout a research project for investigators funded by NSF OCE Biological and Chemical Oceanography Sections or NSF OPP ANT Organisms & Ecosystems Program, with the goal of improving access to NSF funded research data.
OBIS USA	OBIS-USA's role with respect to OA or OA related variables all US sources and applications of biological data: • Mobilize diverse sources of biological occurrence data: Presence-Absence-Abundance • Enable applications and data type integration • Standards: semantics, richness, suitability for applications, discovery, access

	• Infrastructure of Federal Data Lifecycle
CDIAC/DOE	Ocean CO_2 data from oceanographic ships and other platforms • WOCE Database (1991-1999, original data and documentation from all 74 cruises with CO_2 -related measurements) • CLIVAR Repeat Hydrography and Carbon Database (2001-present: WOCE Repeat Sections) • VOS Underway pCO_2 Database (2001 – present) • Moorings and Time Series Database (2003 – present) • Global Coastal Program Data (2005 – present) Data synthesis projects • GLODAP Database (Data synthesis and evaluation, published in 2004) • CARINA Database (Atlantic Ocean data synthesis and evaluation published in 2009) • PACIFICA Database (Pacific Ocean data synthesis and evaluation: in progress, will be published in 2012) • GLODAP-V2 Database (in progress: GLODAP+CARINA+PACIFICA+new Repeat Sections data) • LDEO (Takahashi) Global Surface pCO_2 Database V2010 (first published in 2006, updated every year with new data) • SOCAT (Surface Ocean Carbon Atlas) Database (Published in September 2011, SOCAT-V2 in progress)
OOI	• The OOI science requirements mandate air/sea pCO2, in-water pCO2, CO2 flux, and pH measurements • Appropriate instruments will be installed on surface expression (buoys), water column profilers, and benthic platforms
NASA	NASA's Ocean Biology and Biogeochemistry program focuses on describing, understanding, and predicting the biological and biogeochemical regimes of the upper ocean, as determined by observation of aquatic optical properties using remote sensing data, including those from space, aircraft, and other suborbital platforms.
Ocean Sites	• OceanSITES is an international collaboration to collect & disseminate **open-ocean time-series data**. • Primarily mooring data, but also repeat ship stations. • Data can be of any discipline, but meant to be research-quality. • Organizational structure: - Executive Committee - Steering Team

	- Data Management Team • Data flow structure (after Argo): - Principal Investigator (PI) - Data Assembly Center (DAC) - 2 Global Data Assembly Centers (GDACs: NDBC/Ifremer) • **Ocean acidification data (pH, anything carbon) can be included!** • **CDIAC (A.Kozyr) is on the OceanSITES Data Management Team**

Table 2: Ocean Acidification-relevant Data Streams (identified by workshop participants)

Program Name	**Data Streams**
NOAA NODC	NODC is the designated federal permanent archive for chemical, physical, and biological oceanographic data Underway, CTD/Niskin, Buoys, Plankton, Argo, experimental, model, GTSPP, satellite, glider, Instrumented animals, SeaSor
NOAA Fisheries	Experimental data from response experiments for: commercially important fish and shellfish species, their prey (calcareous plankton) and habitats (corals) Model output for population and socioeconomic consequences forecasts
NOAA Observing Efforts	Underway (SOOP, research cruises and gliders) Moorings
NOAA Ecosystem Modeling	The program integrates across multiple data streams with the goal of assessments of the effect of OA on resources and ecosystems • Climatology (based on in situ data) • Earth System Model Projections • Experimental Results • Population and Ecosystem Models
IOOS	– Observations – Sensors (shore platforms, buoys, gliders, etc) • Mainly physical and chemical variables

	– Water samples (yes, but not IOOS strength) – Modeling – Forecasts, hindcasts – Ocean, weather, ecosystem health – Experimental (very limited) – Satellite – HF Radar network (surface waves)
BCO-DMO	Has special purview over data from all NSF funded projects awarded via the Ocean Acidification special RFPs. Also has data from a wide spectrum of other NSF funded projects. • Observations (from broad-scale and process study cruises, and time-series collection sites) • Profiling, moored, AUV and vessel-mounted sensors • Water sample collectors • Plankton nets, sediment traps • Model Results • Experimental (laboratory and field) • Synthesis Products
OBIS USA	Biological Observations – Human or other basis of observation – Taxon, Coordinates, Date/time – Biological: size, life stage, sex, etc. – Sampling and Observation Method – Quantification, Tracking
CDIAC/DOE	• Observations – Sensors (CO_2 data from VOS, Moorings) – Water samples (CO_2 data from Repeat Section Cruises) • Modeling (Select Ocean Carbon Cycle Model Results Archive) • Experimental (The International Inter-comparison Exercise of Underway fCO_2 Systems During the R/V *Meteor* Cruise 36/1 in the North Atlantic Ocean)

OOI	Observations • Sunburst SAMI pCO_2 and pH instruments • WHOI bulk meteorology package (flux) • Water samples collected during deployment/recovery cruises for calibration/validation purposes Modeling • Some modeling capability through Cyberinfrastructure
NASA	• In situ total alkalinity, pH, pCO2, DIC, PIC, DOC, POC, T, (temporal and spatial scales are cruise dependent) • Global satellite surface winds, SST, salinity, water---leaving radiance products (e.g. chl, calcite), altimetry, scatterometry, modis land products • Modeling (e.g. pCO2, CO2 flux) in regional and global scales
Ocean Sites	• Observations: – Moorings • Surface buoys (e.g. TAO, Papa, Stratus, NTAS, CCE) • Subsurface moorings (e.g. CIS, PAP, ESTOC) • Bottom landers (in progress for MOVE, CORC) – Repeat ship stations • Water samples (e.g. HOT, BATS) • CTD sensors (e.g. HOT, BATS) • Real-time and/or delayed-mode

Table 3. Ocean Acidification-relevant variables (and other information) measured through data collection efforts identified in Tables 1 and 2.

Program Name	**Variables**
NOAA NODC	Measured time scales: 1700s to present (delayed and near real-time). Spatial scales: Global NODC digital archive is variable neutral. Archive contains several collections of biological, chemical, and physical oceanographic data and ocean data products
NOAA Fisheries	"Environmental data": Experimental exposures of organisms to range of elevated CO_2 levels Experimental conditions: Temperature, Salinity, pH, CO_2

<table>
<tr><td></td><td>semi-continuous metering (usually pH)
periodic bottle samples (TA and DIC)
All “artificial” conditions except “ambient” in some experiments

Biological response variables
growth, mortality, metabolism etc
condition-specific
life stage-specific
little standardized terminology
All results specific to experimental conditions</td></tr>
<tr><td>NOAA Observing Efforts</td><td>Carbon Dioxide in water (U M B), Temperature(U M), Salinity(U M B), Oxygen(U M B), pH (U), DIC (U B), Fluorometry (U), Air Temp (M), Total Alklinity (B), Nutrients (B) including nitrite, nitrate, silicate, phosphate

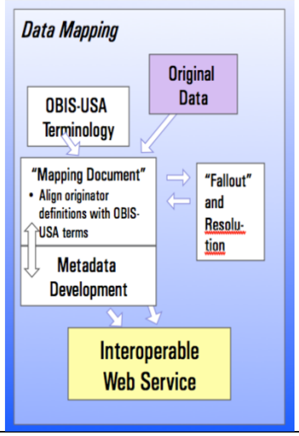
Underway (U), every 3 minutes, duration: weeks-months, no of samples ~6K, total x8=~50k
Moored (M) , every 8 times per day, duration: continuous, no of samples ~11K/year, total x12=~130k/year
Bottle (B), every 3 minutes, duration: 1 time, no of samples ~200-3k, total=~3k</td></tr>
<tr><td>NOAA Ecosystem Modeling</td><td>• Combination of chemical, physical and biological variables
• Spatial scale – 100-1000 kms
• Temporal scale – years to decades
• No standardized vocabularies – working across several disciplines</td></tr>
<tr><td>IOOS</td><td>• Variables
– Focused generally on conventional physical and core chemical variables
– Recent expansion into biological data
– Evolving Water Quality data efforts
• Temporal and spatial scales
– Focused generally (but not exclusively) on recent, real-time, and forecast conditions
– Continuous monitoring (sensors, high-frequency)
– Wide range of spatial scales
• Adopt and support community vocabularies
– CF Standard Names</td></tr>
</table>

	– MMI-hosted vocabularies and vocabulary resources – Extend *only if absolutely necessary*
BCO-DMO	• biogeochemical measurements (OA related) o Carbon cycle chemistry, species data, physical properties, acoustics • Measurements are made for project-specific research themes and contributed by originating investigators (variable temporal and spatial scales) • Names of measurements are mapped to terms from SeaDataNet parameter usage vocabulary hosted by BODC/NERC via SeaVox URL: http://vocab.nerc.ac.uk/
CDIAC/DOE	• Discrete/bottle (DIC, TALK, pH, pCO2, DOC, 14C, 13C, CFCs, other hydrographic data) • Surface/underway (xCO2, pCO2, fCO2 water and air) • Temporal and spatial scales: 1957-Present, Global • Reference for standardized vocabularies: -Discrete: WOCE/CCHDO data format (http://cchdo.ucsd.edu/format.html); - Underway: IOCCP pCO_2 Data File Format developed during the Tsukuba underway data workshop: http://cdiac.ornl.gov/oceans/underway_data_format.html
OOI	• Upper and lower water column pCO2, pH (plus temperature, pressure, salinity, etc.) – pCO2 Units: 0 – 2000 µatm (±2 µatm) – pH range of 7 - 8.5 units (±0.005 units) • Variables measured at all OOI sites (Global, Regional, Coastal) – pH and pCO2 measured within 200 ms of each other, no less than once per hour
OceanSITES	• Mooring measurements typically several times per hour to several times per day • OceanSITES uses netCDF files with CF vocabulary • Many variables already exist, e.g.: sea_water_temperature, sea_water_electrical_conductivity, sea_water_salinity,

	mass_concentration_of_oxygen_in_sea_water, wind_speed, air_pressure_at_sea_level, relative_humidity, concentration_of_chlorophyll_in_sea_water, surface_partial_pressure_of_carbon_dioxide_in_sea_water • Variable names for carbon in the making (pCO_2, pH, TDIC, alkalinity) • Emphasis on metadata (time, location, accuracy, sensor make & model, external references, …)

Table 4: Quality Assurance/Quality Control processes for OA-related data streams, where identified.

Program Name	**QA/QC**
NOAA NODC	Principle Model : NODC OAS follows the Open Archival Information System Reference Model (OAIS) • OAS preserves exact copies (checksum) of all digital data submitted for archival including originator's QC/QA documentation. • If the PI or originator changes the data, then OAS preserves the old and any new versions of the data (Versioning Control) • NODC provides timely unrestricted public access to OA data
NOAA Fisheries	QA-QC based on specific methodologies and lab practices
NOAA Observing Efforts	Measurement protocols published 2007 "Validation cruises" as reliability check for the moored systems Agreed "tests of reasonableness" and standardized QC flags for CO_2 goal: same for O_2, pH, DIC, fluorometry Release within 1 year (CDIAC) Community-based 2nd level QC for CO_2 cruise data (SOCAT, ongoing) mooring data to be added to SOCAT (bottle data : Carina, etc.)
NOAA Ecosystem Modeling	Some kind of review of model code Some kind of review of model set up Are there lessons to be learned from the climate modeling community
IOOS	Revitalizing **QARTOD**, a community effort building consensus toward QA/QC procedures, initially focused on real time data. Through interactions with OA program, QARTOD may be an avenue to work on QC for OA

	parameters in an effort to adopt and support community standards. QARTOD previously focused mainly on physical variables and most conventional chemical variables
BCO-DMO	QA/QC procedures: Done by original PI; documented and available in the data set metadata Metadata are QC'd by BCO-DMO; any changes confirmed with PI Versioning and provenance: Only most recent version is served online Modification history is included in online metadata
OBIS USA	Enrollment: Process, Communication, Technology QA/QC: Assess and Understand, Resolve, Document Update 
CDIAC/DOE	1st level QC/QA: Data, metadata consistency, property-property plots, cruise maps, data vs. time/distance plots for surface measurements, etc. 2nd level QC/QA: – Analytical and calibration techniques – Results of shipboard analysis of certified reference materials – Replicate samples – Consistency of deep carbon data at the locations where cruises cross or overlap – Multiple linear regression analysis – Isopycnal analyses – Internal consistency of multiple carbon measurements – Final evaluation of offsets and determination of correction to be applied http://cdiac.ornl.gov/oceans/glodap/compile.html

OOI	**L0 = Unprocessed Data** – Produced by the instrument – Data is provided in engineering or scientific units – No QA/QC applied **L1 = Basic Data** – Data is provided in scientific units – Some level of instrument calibration and quality control have been performed – QC can be automated or HITL (Human in the Loop) – May include multiple sub-levels (L1a, L1b, etc.) to describe exact processing applied **L2 = Derived Data Products** – Always in scientific units – Always calibrated (meaning that calibrations have been applied, rather than using raw counts; may or may not include 'post-recovery calibration') NASA, NEON, and CODMAC standards were consulted All data levels archived within OOI Cyberinfrastructure QC Algorithims • **Global Range Test**: Generates a QC flag for a data point indicating whether it falls within a given, universally valid range for all applicable data products. • **Local Range Test**: Generates a QC flag for a data point indicating whether it falls within a given range, dependent on the location of the data, hence the name "local". • **Temporal Gradient Test**: Temporally assesses the data streams from multiple instruments on one asset or mooring to generate flags if a data stream deviates significantly from recent data values. • **Spatial Gradient Test**: Spatially assesses the data streams from multiple instruments on one asset or mooring to generate flags if a data stream deviates significantly from data values on nearby instruments. • **Trend Test**: Tests time-series to determine whether the data contain a significant portion of a polynomial. The purpose of this test is to check if a significant fraction of the variability in a time series can be explained by a drift, possibly interpreted as a sensor drift. This drift is assumed to be a polynomial of specified order, e.g., 1 for linear drift. • **Stuck Value Test**: Tests time series for "stuck values", i.e. repeated occurrences of one value, either

	temporally or spatially. • **Spike Test**: Generates flags for data values according to whether a single data value deviates significantly from surrounding data values, either temporally or spatially.
OceanSITES	• Observing platforms are "owned" by a PI, who uses his/her own QA/QC procedures • OceanSITES files contain metadata that are meant to document methodology, accuracy, pass/fail QC flags • OceanSITES as a project is a forum to discuss & establish community best-practices documentation • OceanSITES public file access is via two mirrored GDACs, which contain only the most recent ("best") version of the files. Metadata contain fields for file history and processing.

Table 5: Data archival strategies/requirements

Program Name	Data Archival
NOAA NODC	NODC is the designated federal permanent archive for chemical, physical, and biological oceanographic data
NOAA Fisheries	No systematic archiving in general but NOAA Ocean Acidification Program is requiring NODC archiving
NOAA Observing Efforts	• NOAA's CO_2 obs are archived at CDIAC. From CDIAC they are transferred to NODC. • SOCAT (international) CO_2 partners have nationally affiliated archive centers • OA (multi-agency) will likely need an analogous approach
NOAA Ecosystem Modeling	Model versioning is important; looking for advice on how best to store model versions Model output will need to be stored along with a "set-up" file CMIP5 could be "an example" Code for accessing and analyzing model results should be archived
IOOS	*Facilitate* partnerships that result in robust, long-term archival of data NODC as preferred mechanism for national data archival. NetCDF CF templates will play strong role *Define and monitor* RA "maturity" levels with respect to data archival for regional assets *RA's*: interim / mid-scale archival in a variety of formats and access mechanisms

BCO-DMO	• Data center/repository partners include: GCMD, SeaBASS, CDIAC, OBIS-USA • All data managed by BCO-DMO are submitted to NODC for permanent archive • Data formats vary, but most data are submitted to NODC as plain text ASCII files accompanied by FGDC-compliant metadata and supplemental documentation as appropriate
OBIS USA	Coordination with NODC • OBIS-USA encourage and assist all data contributors to archive entire data product at NODC • Provide referral to NODC as well as joint application opportunities • Entire OBIS-USA resource will be NODC-archived
CDIAC/DOE	• Data management and data archival for ongoing ocean CO_2-related measurement projects and experiments (WOCE, CLIVAR, VOS, Moorings, Coastal and other) • Provide the scientific community and other users with high-quality Ocean CO_2-related original measurements and documentation • Data synthesis involvement – Global and Regional databases (data products) for discrete and surface CO_2-related and other data (GLODAP, CARINA, PACIFICA, LDEO, SOCAT, GLODAP-V2 Databases) • Communication with other oceanographic data centers and projects on data exchange and cooperation (WHPO/CCHDO, OCB/BCO-DMO, NODC, JODC, IOOS, GOOS, Ocean.US, OceanSITES, EuroSITES, CARBOCHANGE, BODC) • NODC serves now as permanent archive for all ocean data served by CDIAC •
OOI	• Data transmitted to shore via satellite or electro-optical cable • All data levels, algorithms, and appropriate metadata will be archived and available • Stored by Cyberinfrastructure team at UC San Diego and made publicly available through a general web interface • Data will also be distributed to archives, including the national data/data products repositories (e.g. National Geophysical Data Center, National Oceanographic Data Center, and National Buoy Data Center).
OceanSITES	Archiving is PI's business, but OceanSITES GDACs are working with NODC for routine archival of GDAC content. Intent is for NODC to have verbatim copy of OceanSITES GDAC, i.e. in OceanSITES data formats.

Table 6. Current and proposed data serving capabilities

Program Name	**Data Serving**
NOAA NODC	Vision for the OA Data Stewardship Framework • Adaptable framework • Robust rich metadata and data history (PIs keep track of data evolution) • Easy to integrate OA data with other relevant measurements (data products) • Archival data granularity • Unique ID (accession/DOI?) + data version control (coordination) • Easy data discovery and access services (NetCDF templates)
NOAA Fisheries	Most of this type of data is now summarized in tables / figures in publications. Goal of increased availability required by the NOAA OAP: 1. make biological response data available for syntheses & modelers 2. provide more detailed data than would be presented in publications
NOAA Observing Efforts	• Search portals at CDIAC (*http://cdiac.ornl.gov/oceans/*) and Pangaea (*http://www.pangaea.de/*) • Download files from same • Visualize and download arbitrary subsets using LAS (*http://ferret.pmel.noaa.gov/SOCAT_cruise_viewer/*) • REST URLs are available Web pages, too, of course (e.g. PMEL's CO_2 Map and Data Viewer *http://www.pmel.noaa.gov/co2/map/index)*
NOAA Ecosystem Modeling	• Not currently serving OA modeling data. • NOAA ESRL Extremes website provides output to climate models – http://www.esrl.noaa.gov/psd/ipcc/extremes/
IOOS	• IOOS assesses, adopts and refines standard data services for data discovery, integration and access • OPeNDAP/THREDDS for gridded data (more recently, evaluating CF feature profiles for discrete data). *Also OGC WCS* • OGC SOS for discrete data; new SWE Common profile matching CF feature profile • OGC WMS for mapped geospatial views

	• Also: – ERDDAP (translation and visualization), regional (custom) services, etc – User Applications facilitating discovery and access to data
BCO-DMO	• Data managed by BCO-DMO are freely available online, and accessible via standard Web browser client • Data can be exported/downloaded as plain text ASCII (CSV, TAB), NetCDF, ODV, Matlab • Geospatial data are available as OGC WMS, WFS and KML
OBIS USA	• Biological Data: – "Download" from browser-oriented web site: discovery, query and exploration. "Data Dashboard" – Web service availability: today, ERDDAP and WMS. Additional web services planned (GeoPortal, others) – Encourage open-source approach to connecting to the resource • Metadata: – Discovery: Metadata Clearinghouse, GCMD, Data.gov. FGDC currently central, ISO-awareness is priority for 2012
CDIAC/DOE	Global Ocean Data Analysis Project (GLODAP) at cdiac.ornl.gov/oceans/ Web-Accessible Visualization and Extraction System (W.A.V.E.S) at cdiac3.ornl.gov/waves
OOI	OOI Cyberinfrastructure (UCSD) will allow users to: • Interact remotely with observatory platforms and instruments • Add or reconfigure sensors on the observatory • Freely and openly access real-time and near real-time data via the Internet • Subscribe to data streams from instruments of interest • Run models within the integrated observatory network • Collaborate with other users in • virtual lab spaces to analyze data, • share ideas and model results
NASA	• No Subscription program. Requires a human to search via web interface and download • Want to implement OPENDap and Thredds, but not authorized or implemented

OceanSITES	• netCDF files on either one of two GDACs available via ftp (modeled after Argo) • Also: THREDDS, OpenDAP • Desire to have more graphical/interactive data selection tool (LAS?) OceanSITES GDAC Top Level Directory on FTP server (ftp://data.ndbc.noaa.gov/data/oceansites) OceanSITES have unique site, platform, and deployment codes.

Appendix 2
Issues from Breakout Discussions Which Need to be Addressed

Issues to resolve for OA data sets in general:

- What is an "OA dataset"? T, S and at least 2 variables of carbonate system (whether measured or derived).
- What constitutes the smallest attributable unit – and how does updating one variable get reflected in the dataset?
- What defines a final product?
- How to distinguish discrete/complete datasets (eg published) from open-ended datasets (eg. Real time data from observing systems)? Should we "chunk" open-ended data sets? Do we assign new DOIs for new versions or can the same DOI apply? If multiple levels of DOIs, how is this tracked?
- Dataset publication options:
    - Are data papers a good option? Journals now evolving for data publication.
    - There is a continuum from data centers (CDIAC doi-only, no publication), Pangea doi + pub doi) to "Dryad" (data files formally linked to traditional journals. Journal linked may be most appropriate for experimental data.
    - It is agreed that data should only have to be submitted ONCE, regardless.
    - Groups already working on this include: CDIAC, BCO-DMO, Pangea, Dryad (bio/ecology), NSF DataOne, SCOR/IODE and IGSN/EarthChem.
    - A project has been underway for several years to address the challenges presented by publication of scientific data. One part of this project involves a recognition of the importance of globally unique identifiers. The project URL: http://www.iode.org/index.php?option=com_content&view=article&id=110&Itemid=129
      the report from the most recent two workshops:
      SCOR/IODE/MBLWHOI Library Workshop Report No. 230
      April 2010 Workshop on Data Publication
      http://www.iode.org/index.php?option=com_oe&task=viewDocumentRecord&docID=5437
      SCOR/IODE/MBLWHOI Library Workshop Report No. 244
      November 2011 Workshop Report:
      http://www.iode.org/index.php?option=com_oe&task=viewDocumentRecord&docID=8098
- What kind of peer-review should be required for a dataset? Would it be possible to provide comments on a dataset for public viewing?

- Need to be able to track provenance of a dataset.

Issues discussed for <u>observing data</u>, specifically.

- From the following platforms: buoys, profiling gliders/AUVs, discrete ship collections, underway/pumped ship, wave glider, cabled sensor, pier/land based continuous and discrete systems, satellite, field experiments
- PI responsibilities: a) define metadata and procedures, b) methods, calibration, precision, accuracy, discoverability, understandability, c) provide both versions: one unadjusted (do we post unadjusted data to web?) and one adjusted within one year of collection.
- Data management experts responsibilities: a) provide set up or auto conversion of data (engineering units) to recognizable units, b) post data to web, c) display graphic and provide access to underlying data, d) provide and document QC for real-time filtering (flags)
- Data Service Center: 1) has responsibility to display data from PIs but PIs can post elsewhere as well, 2) assures long term preservation and attachment of living globally unique identifiers, 3)interact with other data service centers to assure interoperability of data sets. NDBC, PMEL, IOOS or NODC for RT data, CDIAC for non-RT data, NODC archive for all.

Issue to resolve for Laboratory-based and In situ experimental Research

- Public sharing of experimental data is a complicated topic. First, few standards exist for the parameters measured (unlike observing data). Second, experimentalists are not accustomed to sharing data openly before results are analyzed in publication.
- Need for an “in situ” experiments scoping workshop in the US. In situ research is particularly complex because have both observing and experimental data in one project.
- Timeframe requirements for experimental data release must be realistic but…”wait for the paper” is no longer sufficient. One year from original data collection, though, isn’t generally enough time for analysis. Sometimes experiments must be repeated several times over several years before there is enough evidence to publish. Most researchers won’t release data before publication of a paper. However, it may make sense to share data amongst researchers and with data managers (not publicly) prior to publication. Sharing could allow scrutiny and quality assurance/control that wouldn’t have been possible before. If paper hasn’t been written after a reasonable time frame (3 years from original data collection), then data can be made public?

- What will the QA/QC requirements be for experimental data?
- We should consider what the goals are for publicly sharing other than the general requirements that all federally funded data be made public eventually. Modelers definitely need experimental data results. It is important the data be intercomparable for meta-analytical purposes.
- Proposed information to be included in metadata:
    - Experimental set up needs to be well described
    - Were conditions manipulated or use the natural condition
    - Organisms need to be well described.
    - Treatment type: A flat-line or variable (diurnal) system.
    - Does the treatment represent a real place in the world
    - Agreed upon set of experimental or observational qualitative flags
    - Validation sampling process
    - Organism response variables need to be standardized?
    - Develop a core and extended standardized parameter list so data from different researchers can be cross compared.

Appendix 3

List of Participants:
Consortium for the Integrated Management of Ocean Acidification Data

1. Alexander Kozyr, DOE, Oak Ridge National Lab, CDIAC

2. Burke Hales, Oregon State University

3. Chris Sabine, NOAA Pacific Marine Environmental Laboratory

4. Cyndy Chandler, Wood Hole Oceanographic Institution & NSF Biological and Chemical Oceanography Data Management Office

5. David Kline, UC – San Diego/Scripps Institution of Oceanography

6. Emilio Mayorga, University of Washington & NANOOS-IOOS

7. Hernan Garcia, NOAA National Oceanographic Data Center

8. Jan Newton, University of Washington & NANOOS-IOOS

9. Jon Hare, NOAA North East Fisheries Science Center

10. Kevin O’Brien, NOAA Pacific Marine Environmental Laboratory

11. Kimberly Yates, United State Geological Survey

12. Krisa Arzayus, NOAA National Oceanographic Data Center

13. Libby Jewett, NOAA Ocean Acidification Program

14. Libe Washburn, University of California Santa Barbara

15. Liqing Jiang, NOAA National Oceanographic Data Center

16. Michael Vardaro, Oregon State University & Ocean Observations Initiative

17. Mike McCann, Monterey Bay Aquarium Research Institute

18. Paul McElhany, NOAA Northwest Fisheries Science Center

19. Peter Griffith, NASA

20. Philip Goldstein, OBIS-USA

21. Richard Feely, NOAA Pacific Marine Environmental Laboratory

22. Roy Mendelssohn, NOAA Southwest Fisheries Science Center

23. Samantha Siedlecki, University of Washington & JISAO

24. Sean Place, University of South Carolina

25. Simone Alin, NOAA Pacific Marine Environmental Laboratory

26. Steve Hankin, NOAA Pacific Marine Environmental Laboratory

27. Tom Hurst, NOAA National Marine Fisheries Service AFSC

28. Uwe Send, UC – San Diego/Scripps Institution of Oceanography

29. Sarah Cooley (via phone), Woods Hole Oceanographic Institution , Ocean Carbon Biogeochemistry Program

30. Derrick Snowden (via phone), NOAA Integrated Ocean Observing System

31. Jean-Pierre Gattuso (via phone) OA- International Coordination Center

Appendix 4
Proposed Data Management Approach for NOAA Observing Data

Draft March 6, 2012

Contributors: R. Wanninkhof, A. Sutton, S. Alin, C. Cosca, D. Greeley, and R. Feely

Data Submission and Secondary Quality Control/Quality Assurance for Ship of Opportunity, Research Cruises and Mooring Data.

INTRODUCTION

The envisioned ocean acidification (OA) Data Stewardship System (OADSS) will serve a diverse set of OA data from all funded participants in the NOAA OA program in a seamless and transparent manner. However, the submission of data will occur by platform and/or approach. This document outlines submission and metadata requirements for OA data from ships of opportunity (SOOP) and moorings, and Niskin samples and CTD data from research cruises. In addition, this document deals with the validation data for SOOP and moorings that have commonality with the Niskin data in that they are discrete samples taken from Niskin bottles or underway seawater lines on SOOP or from research ships near moorings.

The data submission and QA/QC procedures outlined are designed to facilitate open dissemination of a consistent and high-quality dataset. Moreover, contextual data queries that are using data from different platforms at a similar location or time should be facilitated. We stress the need for acknowledgement and reference to the providers of the data. The guidelines are based on the protocols established in other biogeochemical programs. It is focused on three major program elements of the Ocean Acidification monitoring effort with their associated data streams provided in Table 1:

1) Ship of opportunity efforts: high-frequency surface water data
2) Mooring data: autonomous sensor data
3) Validation data for moorings and SOOP
4) Dedicated research cruises: CTD/Niskin data, high-frequency surface data discussed under SOOP, deck incubations*

*: This data stream is not discussed further here

Table 1. Data types

	SOOP	Mooring	Validation	Cruises
Type	underway	autonomous	discrete	discrete
Parameters	Table 2	Table 5	Table 6	Table 7
Frequency	3-minutes	8-x day	4-x year	1-x year
Duration	weeks-month	continuous	days-week	weeks-month
Typical number of samples	6000	11,000/yr.	20-100	200-3000
Analysis	in situ	in situ	shore side	ship based
Number platforms (Approx.)	8	12	60	2

---

1) SHIP OF OPPORTUNITY UNDERWAY pCO_2

SOOP data submission and metadata

The core SOOP data are similar to the CO_2 measurements in support of sea-air CO_2 flux determinations. It is recommended to take advantage of the procedures set up in these programs that can be augmented to accommodate the data stream for OA.

The data and metadata submission of underway pCO_2 data should follow that recommended at the Carbon Dioxide Information Analysis Center (CDIAC, cdiac.ornl.gov/oceans). The metadata form is shown at:

http://cdiac3.ornl.gov/forms/underwayform.htm

The data submitted should include the core information such that the calculated parameters can be recreated. The core measurements that should be provided are shown in Table 2 along with units.

Table 2 Data fields for underway files

Group_Ship
Cruise ID
JD_GMT
DATE_UTC__ddmmyyyy
TIME_UTC_hh:mm:ss
LAT_dec_degree
LONG_dec_degree
xCO2_EQU_ppm
xCO2_ATM_ppm
xCO2_ATM_interpolated_ppm
PRES_EQU_hPa
PRES_ATM@SSP_hPa
TEMP_EQU_C
SST_C
SAL_permil
fCO2_SW@SST_uatm
fCO2 QC_FLAG
fCO2_ATM_interpolated_uatm
dfCO2_uatm
Comments on flag
Oxygen_umol/kg
O2_ QC_FLAG
pH_sw
pH_ QC_FLAG
DIC_umol/kg
DIC_QC_FLAG
Fluorometry_ug/l
Fluorometry_QC_FLAG

Calculation of fCO_2 should follow the approach for SOCAT and outlined in Pierrot et al. [2009].

Other measurements on OA SOOP:
For ocean acidification monitoring, additional measurements that comprise the core OA suite include oxygen, pH, DIC, and fluorometry, but there is less experience and knowledge of data quality for these data streams. Data quality control procedures should be established for these measurements, including likely ranges. For [surface] oxygen, deviation from saturation, and in cases of large deviation, the anti-correlation with pCO_2 are useful for "reasonability" checks. Validity of pH measurements can be checked by anti-correlation with pCO_2 trends, and/or through calculating pH from pCO_2 and estimates of TA (e.g. from TA-salinity relationships [Lee et al. 2006]).

The additional data are often appended to the same data file, as shown in Table 2. In cases of separate files taken by different investigators they can be readily merged into the data streams within the OADSS , as long as time and location information is provided.

SOOP Quality Control and Quality Assurance (QA/QC)

Primary QC at the individual measurement level:

Submissions to OADSS are expected to have undergone quality control and include the appropriate metadata. Data files should be submitted with quality control flags at the level of individual measurements. For SOOP data this includes checks and flags for outliers beyond reasonable values (based on location, expected range, expected variability or lack thereof) and comparison with climatological values or previous data in the region. Bad data do not need to be submitted unless there appears merit to do so (e.g. some of the other information in the data string has value). The number of bad data should be mentioned in the metadata. Questionable data should be submitted and explained in the metadata. It is recognized that there is a level of subjectivity in the assignment of flags. For SOOP underway pCO_2 data the following flags and subflags (= descriptor of a questionable value is used (**Table 3**):

Table 3 Quality control (QC) flags and subflags used in data reduction of underway pCO2 data

QC_FLAG: Quality control flag
2 = Good value
3 = Questionable value
4 = Bad value

QC_SUBFLAG: Descriptive quality control flag used when a value receives a "3" QC flag
1 = Outside of Standard Range
2 = Questionable/interpolated SST
3 = Questionable EQU temperature
4 = Anomalous ΔT (EqT – SST)(± 1°C)
5 = Questionable Sea Surface Salinity
6 = Questionable pressure
7 = Low EQU gas flow
8 = Questionable air value
9= Interpolated standard value
10 = Other, see metadata

Secondary QC:
The secondary QC is performed at OADDS and should be done in consultation with the investigators. Data received by OADSS should undergo automated range and reasonable checks to assure that submitted data have undergone a reasonable level of primary QC. Issues with incorrect column headings, units, and default values are often discovered and easily corrected at this stage. An automated comparison with climatological data or other data in the OADDS or related NODC databases should be performed. Any anomalous patterns and possibly questionable values should be discussed with the investigator who submitted the data. Recommended flagging of individual points as well as flags for whole dataset provided by OADDS should be confirmed with the investigator submitting the data. Reasons for not adhering to OADDS recommendations should be described in metadata. While the OADDS staff will make the recommendations, it is recognized that the investigators have a higher level of expertise and should be actively engaged. Allocating funds for regional groups to engage in secondary quality control, as is done in SOCAT, is a cost-effective and efficient way to engage experts in the process.

*Quality-control flags at the whole dataset level***:**
Following the example of the Surface Ocean CO_2 Atlas SOCAT (http://www.socat.info/), the submitted SOOP, and validation data submitted to OADSS should receive a quality flag ranging from A-F after secondary QC. The explanation is provided in **Table 4.**

Table 4. Quality control flag for full datasets

A: Follows the best practices sampling and analyses procedures and data are deemed good
B: Follows most standard sampling and analyses procedures and data are deemed good
C: Does not follow standard sampling and analyses procedures but data are deemed good
D: Follows most standard sampling and analyses procedures but data are questionable
F: Does not follow standard sampling and analyses procedures and data are questionable

or bad
----

2) MOORING OA DATA

Mooring Data submission and Metadata

High-resolution oceanic and atmospheric pCO_2 time-series data and metadata from moorings are currently archived at CDIAC and incorporated into a variety of synthesis and modeling projects with the overall goal of better understanding the oceans role in climate and climatic change . It is recommended to take advantage of the procedures set up via CDIAC that can be augmented to accommodate the data stream for additional OA parameters.

The data and metadata submission of mooring pCO_2 data should follow that recommended for underway pCO_2 at CDIAC. The metadata form is shown at: http://cdiac3.ornl.gov/forms/underwayform.htm

The data submitted should include the core information such that the calculated parameters can be recreated. The core measurements that are currently provided are shown in **Table 5** along with units and a brief description.

Table 5. Current data fields for CO_2 mooring files
Mooring name
Latitude
Longitude
Date UTC - format: mm/dd/yyyy
Time UTC - format: hh:mm
xCO2_SW_wet [μmol/mol] - mol-fraction of CO_2 in sea water in wet gas
xCO2_SW_wet - QF Quality Flag
H2O_SW [mmol/mol] - mol-fraction of H_2O in sea water
xCO2_Air_wet [μmol/mol] - mol-fraction of CO_2 in air in wet gas
xCO2_Air_wet_QF - Quality Flag
H2O_Air [mmol/mol] - mol-fraction of H_2O in air
Licor_Atm_Pressure [hPa] - Atmospheric Pressure
Licor_Temp [Deg. C] - Atmospheric Temperature
%_O2 – O2 measurement made in equilibrated air (not a quantitative measurement)
SST [Deg. C] - Sea Surface Temperature
SSS - Sea Surface Salinity
xCO2_SW_dry [μmol/mol] - mol-fraction of CO_2 in sea water in dry gas
xCO2_Air_dry [μmol/mol] - mol-fraction of CO_2 in air in dry gas
fCO2_SW_sat [μatm] - Fugacity of CO_2 in sea water
fCO2_Air_sat [μatm] - Fugacity of CO_2 in air
dfCO2 [μatm] - difference (fCO2_SW - fCO2_Air)

Calculation of fCO_2 should follow the approach for SOCAT and outlined in Pierrot et al. [2009].

Other Measurements on OA Moorings:
Data quality control procedures should be established for additional surface and subsurface measurements that comprise the core OA suite such as dissolved oxygen, pH, fluorescence, and turbidity as discussed in the SOOP section. Data fields will need to be added to **Table 5** for these parameters in addition to associated QF quality flags and essential diagnostic information. CDIAC is supportive of adding these measurements to their CO_2 mooring data archive.

Mooring Quality Control and Quality Assurance (QA/QC)

*Mooring QC Flags***:**
Mooring QC flags follow a slightly adapted WOCE QC protocol with main designations of "1"= preliminary data, no QC; "2"= acceptable measurement; "3" = questionable measurement; "4" = bad measurement; "5" = not reported.

Submissions to CDIAC are expected to have undergone quality control and include the appropriate metadata. Data files should be submitted with quality control flags. For mooring data this includes checks and flags for internal CO_2 calibration, for problems with CO2 system diagnostics, for outliers beyond reasonable values (based on location, expected range, expected variability or lack thereof), and comparison with climatological values or previous data in the region. For example, seawater and air CO_2 in each data set is compared with the Marine Boundary Layer data from GlobalView-CO2 and corrected accordingly. Adjustments are also made to the Licor sensor pressure is also made based on each sensor's bias to barometric pressure as measured in the lab.

Bad data are often submitted as there is usually other information in the data string of value. Questionable data should be submitted and explained in the metadata. It is recognized that there is a level of subjectivity in the assignment of flags. For additional information, see Sabine [2005].

3) VALIDATION SAMPLES: SOOP AND MOORINGS

Validation samples are desired at a frequency of 4 times a year or more. The discrete data are either taken from the underway line feeding the pCO_2 system or taken from Niskin bottles and include depth samples. The term "validation" is used in its broadest sense in that the samples are used to verify the continuous data, and for analysis of carbon parameters that cannot currently be done autonomously such as DIC and TA. They can include subsurface data. These samples along with the SOOP and mooring data are critical to establish an "OA product suite". [see e.g. Gledhill et al. 2009; Juranek et al., 2009]

The files submitted to the cognizant OA data management office should have a format and metadata format similar to the research cruise data (see Table 6).

Table 6. Validation samples (Discrete)

Column headers for data from validation samples with units:

Group
Cruise_ID
Station
Sample_ID
Date (UTC)
Time (UTC)
JD_sample
Latitude (decimal +=N)
Longitude (decimal + =E)
Pressure (db)
Salinity_Discrete
Salinity_Continuous
Temperature (deg C)
DIC (umol/kg)
QC_Flag_DIC
fCO_2_Discrete
QC_Flag_fCO_2_Discrete
TAlk (umol/kg)
QC_Flag_TAlk
Oxygen_Discrete

QC_Flag_Oxygen_Discrete
Oxygen_Continuous
NO2 (umol/kg)
QC_Flag_NO2
NO3 (umol/kg)
QC_Flag_NO3
SIO3 (umol/kg)
QC_Flag_SIO3
PO4 (umol/kg)
QC_Flag_PO4
Additional parameters
QC-flags additional parameters

QC-flags will follow a slightly adapted WOCE QC protocol with main designations of "1"= preliminary data, no QC; "2"= good; "3" = questionable; "4" = bad; "6" = duplicate
----

Sampling and Analysis of Validation samples:

These bottle samples should be sampled according to the best practices protocols [DOE, 1994; Dickson et al. 2007, 2010; Riebesell et al., 2010]. At minimum they should analyzed for salinity DIC and TAlk. Following best practices protocols and the need to take several samples from a single bottle, 500-ml high-density borosilicate glass sample bottles should be used.

Duplication:
Approximately 20 % of the bottle samples (1 in 5) should be done in duplicate to assess precision of samples. For an assessment of overall sample integrity duplicate sample bottles should be taken consecutively from the underway-sampling line or from a single Niskin taken at depth. Additional information from sensors should be logged and carried through in the submitted data (T, S, depth, location, time stamp). The bottle samples used for validation should be provided as a separate data file (from the underway or buoy data)

Meta Data for Validation Samples:
The metadata should include the information as outlined in
http://cdiac3.ornl.gov/forms/discreteform.htm
While this form is specific for inorganic carbon data, it can be easily augmented with the other parameters sampled.

4) RESEARCH CRUISES

The shipboard cruise data sets generally have two formats. The first is for underway measurements which follow the same formats for data and metadata submission as outlined in section 1. The second is for discrete samples collected from the CTD/Rosette/Niskin bottle samples. The samples are collected according to the according to the best practices protocols [DOE, 1994; Dickson et al. 2007, 2010; Riebesell et al., 2010] as described in the Repeat Hydrography data reports [for example, see Feely et al., 2009].

Before submission, the discrete data will have gone through its preliminary (primary) stage at the institutions responsible for discrete measurements. During this time the data should be carefully examined by the responsible PI for possible obvious outliers and internal consistency, the corrections for post cruise calibrations are applied, and the quality flags for the carbon data are assigned. Also, adjustments based on shipboard CRM analyses are made. As soon as the discrete data from the repeat sections are transferred to CDIAC as Secondary Research Data, CDIAC will release the data to public via the WWW Live Access Server (LAS) with a Secondary Research Data tag and perform the basic QA/QC Received carbon-related data will be merged with the final hydrographic measurements and the file will be put in the uniform format. If the hydrographic and chemical data are not available at the time of receiving the carbon measurements, CDIAC will contact the cognizant group and work together in preparation of the final data set. If problems with data are found during the QA-QC, CDIAC will contact the responsible PI(s). PI, CDIAC, and OADDS the will work together to resolve all problems in order to upgrade the data to archival data status. Approximately 10% of the bottle samples should be done in duplicate to assess precision of samples. Additional information from sensors should be logged and carried through in the submitted data (T, S, depth, location, time stamp) along with the bottle data. The files submitted to the OADDS should have a format similar to the format given in **Table 7** below.

Table 7. Cruise Niskin bottle data showing column headers with units:

Group
Cruise_ID
Station
Sample_ID
Date (UTC)
Time (UTC)
JD_sample
Latitude (decimal +=N)
Longitude (decimal + =E)
Pressure (db)
Salinity_Discrete
Salinity_Continuous
Temperature (deg C)
DIC (umol/kg)
QC_Flag_DIC
fCO_2_Discrete
QC_Flag_fCO_2_Discrete
TAlk (umol/kg)
QC_Flag_TAlk
Oxygen_Discrete
QC_Flag_Oxygen_Discrete
Oxygen_Continuous
NO2 (umol/kg)
QC_Flag_NO2
NO3 (umol/kg)
QC_Flag_NO3
SIO3 (umol/kg)
QC_Flag_SIO3
PO4 (umol/kg)
QC_Flag_PO4
Additional parameters
QC-flags additional parameters

QC-flags will follow a slightly adapted WOCE QC protocol with main designations of
“1”= preliminary data, no QC; “2”= good; “3” = questionable; “4” = bad; “6” = duplicate
----

The metadata should be submitted along with the data file in a format similar to Figure 1 below.

Figure 1. Example of shipboard cruise metadata for discrete bottle data.

Figure 1. Continued.

Acknowledgement and Attribution of Data

A critical issue is that there is limited reward for an investigator to submit data in a timely fashion. Invariably performance is rated on publications and citations that are related to the datasets obtained. To increase needed recognition for the data gatherers, the datasets must remain clearly linked to the investigator even when incorporated in an inclusive and/or query based data holding. A clear data acknowledgement protocol needs to be established and available to all investigators who use the data. Users of the data should be encouraged to consult with investigators who submitted the data regarding quality, patterns, and interpretation. The NOAA OAP should be acknowledged in any publications and a record of publications using the OA data should be maintained and served by the OADDS.

References

Dickson, A.G., Sabine, C.L., Christian, J.R., 2007. Guide to best practices for ocean CO_2 measurements. PICES Special Publication 3, 191 pp. (see, *http://cdiac.ornl.gov/oceans/Handbook_2007.html*)

Dickson, A.G., 2010. Part 1: Seawater carbonate chemistry In: Riebesell U., Fabry V. J., Hansson L. & Gattuso J.-P. (Eds.), Guide to best practices for ocean acidification research and data reporting, 260 p. Publications Office of the European Union., Luxembourg. (see, http://www.epoca-project.eu/)

DOE 1994. Handbook of methods for the analysis of the various parameters of the carbon dioxide system in sea water; Dickson and Goyet eds. version 2. DOE. (see, *http://cdiac.ornl.gov/oceans/handbook.html*)

Feely, R.A., C.L. Sabine, F.J. Millero, C. Langdon, A.G. Dickson, R.A. Fine, J.L. Bullister, D.A. Hansell, C.A. Carlson, B.M. Sloyan, A.P. McNichol, R.M. Key, R.H. Byrne, and R. Wanninkhof (2009): Carbon dioxide, hydrographic, and chemical data obtained during the R/Vs Roger Revelle and Thomas Thompson repeat hydrography cruises in the Pacific Ocean: CLIVAR CO_2 sections P16S_2005 (6 January–19 February, 2005) and P16N_2006 (13 February–30 March, 2006). ORNL/CDIAC-155, NDP-090, A. Kozyr (ed.), Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN, 56 pp.

Lee, K., Tong, L.T., Millero, F.J., Sabine, C.L., Dickson, A.G., Goyet, C., Park, G.-H., Wanninkhof, R., Feely, R.A., Key, R.M., 2006. Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans Geophys. Res. Let. 33, L19605, doi: 10.11029/2006GL027207.

Pierrot, D., Neil, C., Sullivan, K., Castle, R., Wanninkhof, R., Lueger, H., Johannson, T., Olsen, A., Feely, R.A., Cosca, C.E., 2009. Recommendations for autonomous underway pCO_2 measuring systems and data reduction routines. Deep -Sea Res II 56, 512-522.

Gledhill, D.K., Wanninkhof, R., Eakin, C.M., 2009. Observing Ocean Acidification from Space. Oceanography 22 (4), 48-59.

Juranek, L.W., R. A. Feely, W. T. Peterson, S. R. Alin, B. Hales, K. Lee, C. L. Sabine, Peterson, J., 2009. A novel method for determination of aragonite saturation state on the continental shelf of central Oregon using multi-parameter relationships with hydrographic data. Geophys. Res. Let. 36, doi:10.1029/2009GL040778.

Riebesell U., Fabry V. J., Hansson L. & Gattuso J.-P. (Eds.), 2010. Guide to best practices for ocean acidification research and data reporting, 260 p. Luxembourg: Publications Office of the

European Union. (see, *http://www.epoca-project.eu/)*
Sabine, C. (2005): High-resolution ocean and atmosphere pCO_2 time-series measurements. The State of the Ocean and the Ocean Observing System for Climate, Annual Report, Fiscal Year 2004, NOAA/OGP/Office of Climate Observation, Section 3.32a, 246–253.