

Data Management for the Long Run: OA needs for today and tomorrow

Krisa M. Arzayus, NODC

“The success of the National Ocean Acidification Enterprise will depend critically on effective data management and integration.”

- The *submitted* Interagency
Working Group Strategic Research Plan on Ocean Acidification.

Outline

- International collaboration
- OA Data Management Workshop (Seattle, 2012)
- Interagency collaboration
- National Mandates and Drivers
- Challenges and Opportunities

International Coordination Centre

- **Ocean Acidification International Coordination Centre (OA-ICC):** announced in June 2012 at the Rio+20 UN Conference; operational since July 2012
- Operated by the International Atomic Energy Agency Environment Laboratories
- Global Ocean Acidification Observing Network

International Data Management

- **OA-ICC Data Management Goals:**
 - Facilitate actions to develop common international data formats, coordinate existing efforts and promote data access and sharing within the ocean acidification research community.
 - Building off of success of EPOCA and EUR-OCEANS
 - Compile published data on the impacts of ocean acidification and make them openly accessible. Currently, this is done via [Pangea](#).
- **Joint OSPAR/ICES OA Study Group (SGOA):** Plans for a workshop on biological response to OA, defining ICES data reporting requirements; next meeting is in October 2013

Global Ocean Acidification Observing Network

Vision proposed at the 2nd GOAON (St. Andrews, 2013): Effective long-term scientific data management using interactive, interoperable, and consolidated online data services (human- and machine-to-machine).

- *Data sharing time limits* for coastal and open ocean
- *Web data portal:* data discovery, access, integration, and visualization from collection- to granular- level OA data and data products (common inter-operable web data services).
- *Coordinated scientific data management* and data flow by building on existing infrastructure and scientific requirements over the long-term (OA-ICC, others).
- *Good documentation and metadata procedures* and quality following international format standards to facilitate data discovery, use of DOIs or similar, clear provenance (attribution).
- International OA long-term archival center for OA observational, biological, model data, data products (provide data integration where possible).

OA Data Management Workshop

- First OA data management workshop held March 2012 in Seattle, WA.
- 1st OA Data Management [Plan](#), June, 2012.

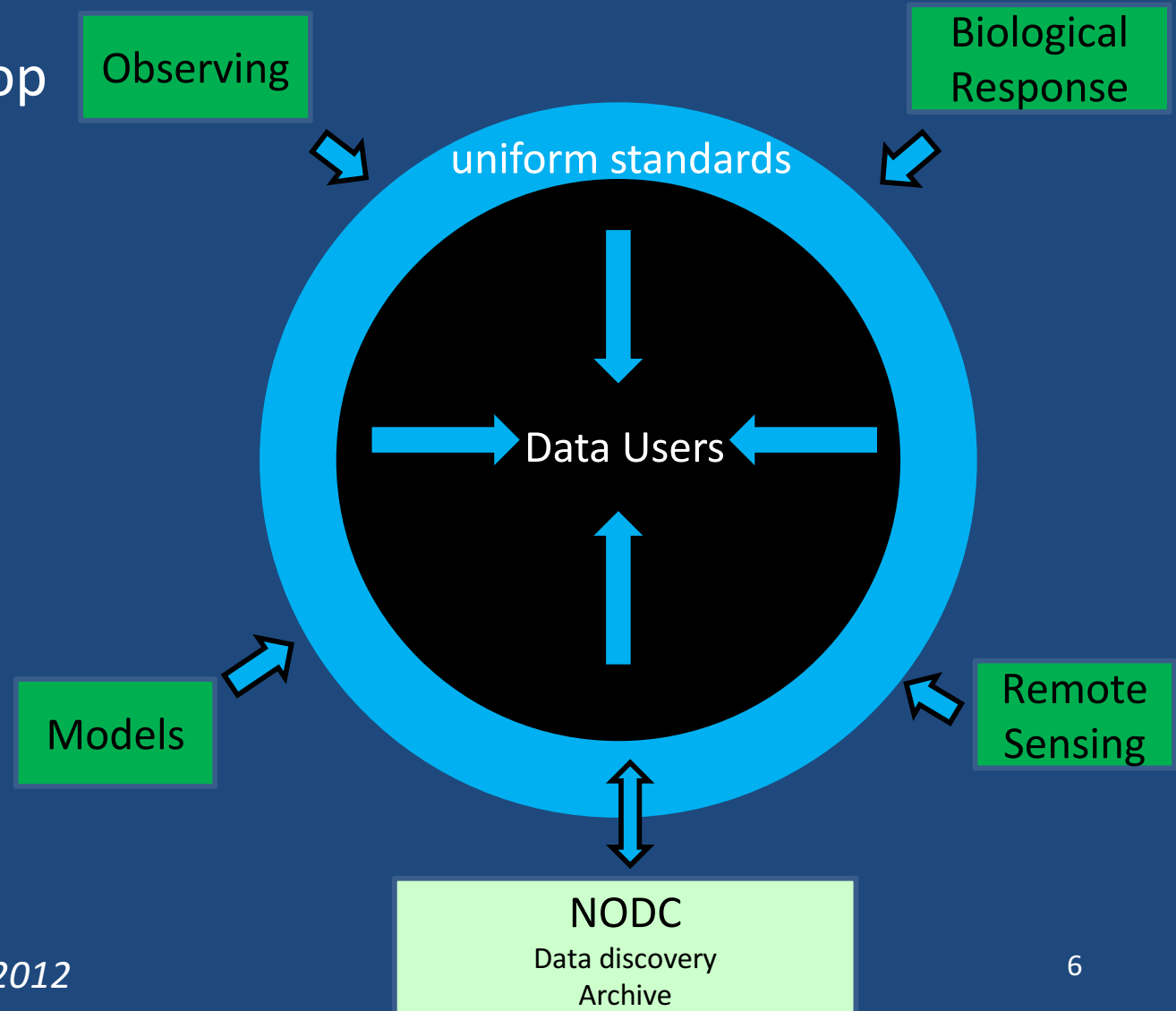


Figure adapted from Arzayus *et al.*, 2012

**US Interagency
Ocean
Acidification Data
Management
Plan: Draft One
June 23, 2012**

US OA Data
Management
workshop in
Seattle, WA on
13-15 March 2012

“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.

Declaration of Interdependence

- **AGENCY SUPPORT:** Top-down, agency directed endorsement of machine-to-machine cataloging and data retrieval protocols to provide synergistic, consolidated mechanisms for scientists to locate and acquire oceanographic data;
- **COMMUNITY ORGANIZATION:** 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
- **SCIENTIST AND DATA MANAGER COLLABORATION:** 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.

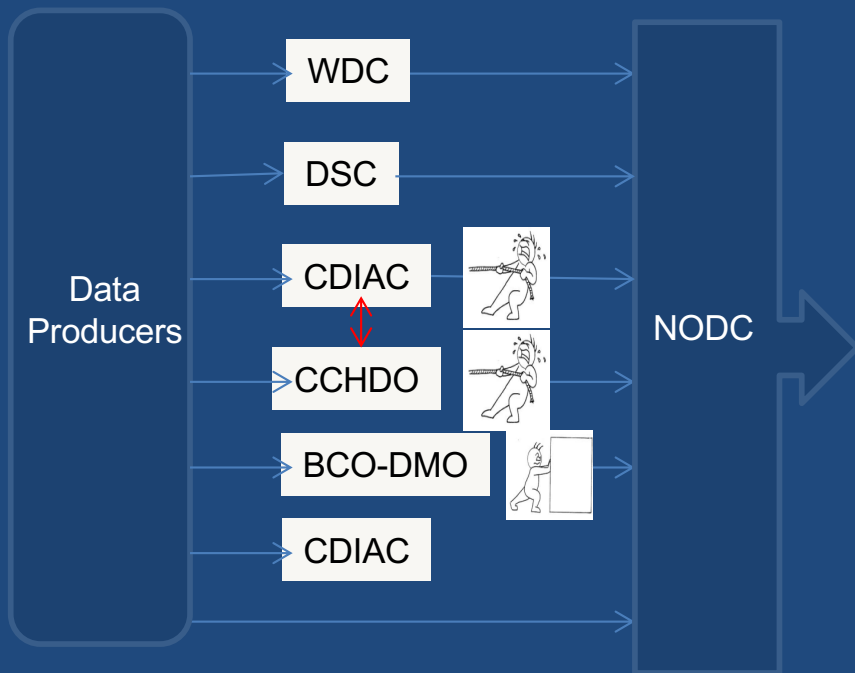
Interagency Collaboration

CDIAC
EPA
OAP
USGS
NODC
NASA
BCO-DMO
OBIS-USA
NSF
OOI
IOOS
FWS
BOEM
CCHDO
NOAA
EarthCube

Data Flow

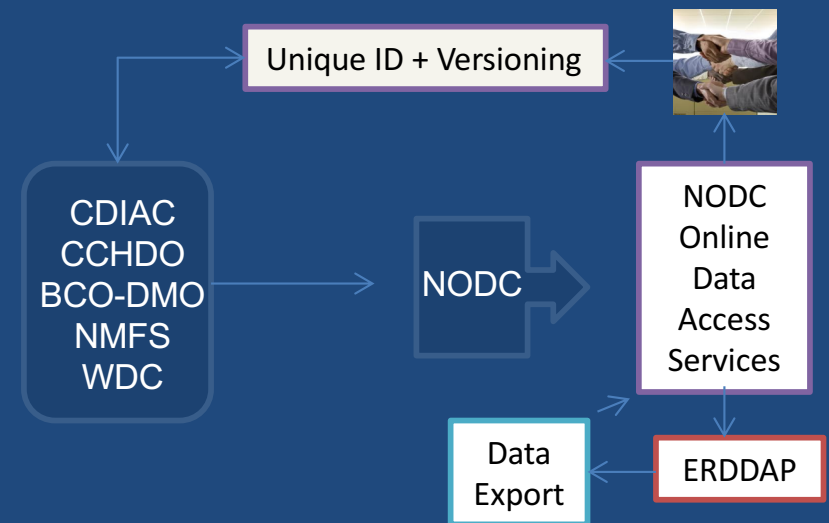
Present

Improvements needed in data flow, data versioning coordination, and data integration



Propose

Coordinated data flow



NODC OADS Automated Retrieval of CDIAC data

A total of 434 Ocean Carbon related data sets, and 3 data products have been transferred to NODC from CDIAC (411 Accessions)

Voluntary Observing Ships

(91 Accessions)

Global Coastal Program

(10 Accessions)

Time Series and Mooring

(27 Accessions)

CLIVAR

(120 Accessions)

PACIFICA

(163 Accessions)

LDEO Data Product

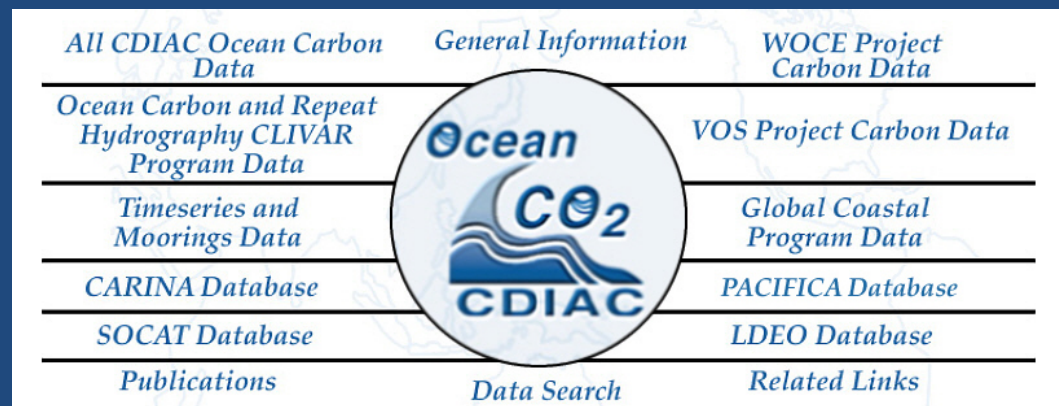
(Accession 0059946)

SOCAT Data Product

(Accession 0101726)

PACIFICA Data Product

(Accession 0110865)



National Mandates and Drivers

- Three Administration-level directives issued in less than a year!
- Bottom line: federally funded research and data collection efforts need to be accessible.

- [Making Open and Machine Readable the New Default for Government Information](#), **Executive Order** dated May 9, 2013
- [Open Data Policy - Managing Information as an Asset](#), **OMB M-13-13** memo dated May 9, 2013.
- [Increasing Access to the Results of Federally Funded Scientific Research](#), **OSTP memo** dated Feb 22, 2013.

Agency and program specific data management requirements are not coordinated

Agency and program resources are not aligned toward common data management objectives

Challenges . . .

- Obstacles to sharing data
- Lack of timely and standardized QA/QC
- Highly distributed OA data are challenging to discover from one place
- Insufficient resources dedicated to documenting and archiving OA data
- Institutional boundaries and limitations
- Competing and growing data management requirements

. . . and Opportunities

- **DOIs:** ensuring data providers get credit for their investment
- **Web services:** providing increasing capability for interoperability
- **Broad data integration initiatives:** GEO-IDE UAF, EarthCube, ICES data requirements, Big Data are fostering interagency collaboration on standards