A metadata template for ocean acidification data

Li-Qing Jiang[1,], Sarah A. O'Connor[2], Krisa M. Arzayus[3], Alex Kozyr[4], A. Rost Parsons[3], James Reagan[1]*

[1] Cooperative Institute for Climate and Satellites-Maryland, University of Maryland, College Park, Maryland USA
[2] ADNET Systems, Inc., Bethesda, Maryland , USA
[3] National Oceanographic Data Center, National Oceanic and Atmospheric Administration, Silver Spring, Maryland, USA
[4] Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

[*] Corresponding author: Liqing.Jiang@noaa.gov

① Abstract

This poster defines best practices for documenting ocean acidification (OA) metadata and presents a framework for an OA metadata template. Metadata is structured information that describes and locates an information resource. It is the key to ensuring that a data set will survive and continue to be accessible into the future. With the rapid expansion of studies on biological response of organisms to OA, the lack of a common metadata template to document the resulting data poses a significant hindrance to effective OA data management efforts. In this poster, we present a metadata template that can be applied to a broad spectrum of OA studies, including those studying the biological response of organisms to OA. The "variable metadata section", which includes the variable name, observation type, whether the variable is a manipulation condition or response variable, and the biological subject on which the variable is studied, forms the core of this metadata template. Additional metadata elements, such as principal investigators, temporal and spatial coverage, platforms for the sampling, data citation, are essential components to complete the template. We also explain the structure of the template, and define many metadata elements that may be unfamiliar to researchers.

② Metadata: definitions and its roles in data management

❑ Metadata is structured information that describes an information resource (e.g., an oceanographic data set). It documents such information as: what was measured; by whom; when (temporal coverage), where (geographic coverage), and how it was sampled and analyzed; with what instruments, and following what protocol; and finally its units of measure and quality of the data.

❑ Metadata helps to document information about the data sets in consistent and standard ways, so that the data sets can be understood and utilized beyond the original use for the data.

❑ Metadata is critical to data discovery by enabling the data sets to be found through relevant criteria, bringing similar data sets together, distinguishing dissimilar data sets, and giving information about locations of the data sets.

❑ Metadata plays an extremely important role in supporting archival and preservation of data, facilitating interoperability, and integrating legacy data. It serves as the "key" to ensuring that a data set will survive and continue to be accessible by future researchers (Guenther, R. and Radebaugh 2004).

③ The need for an OA metadata template

❑ Based on statistics from the European Project on Ocean Acidification (EPOCA)'s bibliographic database, the number of publications on OA averaged about 10-20 per year from 1990 to 2005, then increased sharply to about 270 publications per year by 2011 (Laffoley and Baxter, 2012).

❑ Much of this increase was due to studies on biological response of organisms to OA. As an example, publications on this type of studies accounted for over 80% of all the published OA papers in 2011 (Laffoley and Baxter, 2012).

❑ With the rapid growth in research on biological response of organisms to OA and the parallel rise in corresponding publications, the need for a comprehensive OA metadata template to facilitate archival and access to this important body of data is increasingly evident.

④ Requirements for the OA metadata template

❑ One of the main goals of the metadata template is to enable data discovery.

❑ Another important role of the metadata template is to document as much useful information of a data set as consistently as possible.

❑ OA covers a wide range of oceanographic subject areas, including water chemistry studies, biological monitoring, physiological response experiments, model studies, and paleoceanography studies. If the metadata template can be constructed to apply to many types of OA data sets, the OA data management effort will be much more effective.

⑤ Development process

The development of the OA metadata template involved two steps:

❑ Content standard development to choose the metadata elements that should be included in the template.

❑ Format standard development to assure placement of the metadata elements (from the content standard development) into their appropriate ISO 19115-2 Extensible Markup Language (XML) fields (International Organization for Standardization, 2009; Mize et al., 2011).

⑥ Variable metadata section

The term "variables" (or parameters) refers to the observed or derived properties of a study (e.g., temperature, salinity, dissolved oxygen (DO), chlorophyll, and larval survival rate). Variables are treated as the focal point of the entire metadata template, and all of its child elements are organized around the variable itself.

Table 1. Variable metadata section with child metadata elements organized around the variable/parameter

Root element	Child elements
Variable/ parameter	Variable abbreviation in data files
	Full variable name
	Climate and Forecast standard name
	Observation type
	In-situ / manipulation / response variable
	Variable unit
	Measured or calculated
	Sampling instrument
	Analyzing instrument
	Duration (for settlement/colonization studies)
	Detailed sampling and analyzing information
	Field replicate information
	Uncertainty
	Data quality flag description
	Method reference (citation)
	Biological subject
	Species Identification ID (if available)
	Researcher who measured this parameter — Name, Institution

Among the child metadata elements, **observation type**, whether the variable is a **manipulation condition or a response variable**, and the **biological subject** on which the variable is studied form the skeleton structure of the variable metadata section.

"**Observation type**" identifies the way a variable was captured in relation to its observation context. It could be generic terms that describe how a variable is collected. For example, for water chemistry studies, the observation type could be "surface underway", "time series", or "profile". For physiological response OA studies, such terms as "laboratory experiment", "pelagic mesocosm", "benthic mesocosm", or "natural perturbation site study" could be used (Table 2).

In physiological response studies, variables fall into several categories, some variables like pH, pCO_2, etc. could be **manipulated**, other variables like growth rate, calcification rates are the variables whose **response** to the changing pH is the focus of the study.

In biological studies, many of the measured variables are attached to a specific organism or a biological community. For example, the variable "larval survival rate" is not detailed enough without mentioning either the specific organism or the biological community on which the larval survival rate was studied. We call this "**biological subject**".

⑦ Observation types

Table 2. Selected commonly used observation types of a variable in ocean acidification studies.

Observation types	Definitions	Examples
Surface underway	A series of data points along a path at the surface of a water body with monotonically increasing times.	Surface water pCO_2 measured from a voluntary observing ship.
Time series	A series of data points at the same geographic location with monotonically increasing times.	Water temperature measured on a moored buoy.
Profile	An ordered set of data points along a vertical line (from surface to a certain depth) at a fixed geographic location and fixed time.	Temperature measured from a CTD cast.
Laboratory experiment	Perturbation experiments in enclosed systems (e.g., aquariums) with natural or modified assemblages under modified environmental conditions.	A study the OA effect on shell calcification rate of a species in an indoor aquarium, by bubbling CO_2 to lower acidity of the water.
Pelagic Mesocosm	A mesocosm study has the advantage over standard laboratory experiments in that it maintains a natural community under close to natural, self-sustaining conditions, taking into account relevant aspects from "the real world" such as indirect effects, biological compensation and recovery, and ecosystem resilience (Riebesell et al., 2010). "Pelagic" zone is defined as any water in a sea or a lake that is neither close to the bottom nor near the shore.	(none)

⑧ Other metadata elements

In addition to the variable metadata section, other metadata elements that are included in this template include:

❑ Principal Investigators
❑ Data Producer
❑ Title
❑ Abstract
❑ Purpose
❑ Temporal coverage
❑ Spatial coverage
❑ Geographic names
❑ Platforms (e.g., research vessels).
❑ EXPOCODE, Cruise ID, Section name.
❑ Data citation
❑ Publications describing the data set
❑ Funding agency + Grant no.
❑ Supplementary information

⑨ Template files and availability

The OA metadata template, consisting of three files, can be found at: http://www.nodc.noaa.gov/oceanacidification
❑ a submission form (in Excel spreadsheet) that can be used by Data Producers to prepare their metadata;
❑ an instruction file (in pdf format) that lists the hierarchical relationships of all the metadata elements and their definitions;
❑ an XML file (encoded with the ISO 19115-2 standards) that can be fed into data search portals for OA data discovery.

An older version of the template is also archived in NOAA's Institutional Repository.
❑ http://ezid.cdlib.org/id/doi:10.7289/V5C24TCK.
❑ DOI: 10.7289/V5C24TCK.
❑ NOAA Institutional Repository Accession number: ocn881471371.

⑩ Acknowledgements

❑ Funding support from NOAA's Ocean Acidification Program.
❑ Hernan E. Garcia (NODC), Philip Goldstein (U. Co.) contributed to the content standard development
❑ Jaci Mize (NCDDC) and Sheri Philips (NODC) offered tremendous help on the use ISO 19115-2.
❑ We thank Dean Perry and Dylan Redman for allowing us to use their metadata as a real world example for our template.
❑ The template benefited from discussion with Chris Chambers, Thomas Hurst, Chris Long, Annette DesRochers, Molly Timmers, Nina Bednaršek, Derek Manzello, Renee Carlton, and Jessica Morgan (NOAA), as well as David Kline (Scripps Institution of Oceanography).