



A metadata template for ocean acidification data

L.-Q. Jiang¹, S. A. O'Connor², K. M. Arzayus³, and A. R. Parsons³

¹Cooperative Institute for Climate and Satellites – Maryland, Earth System Science Interdisciplinary Center, University of Maryland, 5825 University Research Ct., College Park, MD 20740, USA

²ADNET Systems, Inc., 6720B Rockledge Dr., Bethesda, MD 20817, USA

³National Centers for Environmental Information, National Oceanic and Atmospheric Administration, 1315 East–West Highway, Silver Spring, MD 20910, USA

Correspondence to: L.-Q. Jiang (liqing.jiang@noaa.gov)

Received: 18 November 2014 – Published in Earth Syst. Sci. Data Discuss.: 13 January 2015

Revised: 11 May 2015 – Accepted: 18 May 2015 – Published: 11 June 2015

Abstract. This paper defines the best practices for documenting ocean acidification (OA) data 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 be accessible into the future. With the rapid expansion of studies on biological responses 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 paper, we present a metadata template that can be applied to a broad spectrum of OA studies, including those studying the biological responses 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 investigators, temporal and spatial coverage, and data citation, are essential components to complete the template. We explain the structure of the template, and define many metadata elements that may be unfamiliar to researchers.

Template availability

- Available at
<http://ezid.cdlib.org/id/doi:10.7289/V5C24TCK>
- doi:10.7289/V5C24TCK
- NOAA Institutional Repository Accession number:
ocn881471371

1 Introduction

Since the start of the Industrial Revolution, human activities have released large amounts of carbon dioxide (CO₂), causing atmospheric CO₂ to increase over 40% compared to the highest CO₂ levels in the last 800 000 years (Hönisch et al., 2009). However, only about half of the emitted CO₂ stays in the atmosphere. The remaining portion has been taken up by the ocean and terrestrial systems. Based on the most recent data, from 2004 to 2013, the global oceans have absorbed

about 26 % of anthropogenic CO₂ emissions (Le Quéré et al., 2014). While effectively slowing the increase of atmospheric CO₂ and hence global climate change, the oceanic uptake of CO₂ is causing decreasing pH, carbonate ions (CO₃²⁻) concentration, and carbonate saturation state. The process is commonly called ocean acidification (OA) (Caldeira and Wickett, 2003; Feely et al., 2004; Orr et al., 2005).

Over the past 10 years, OA studies have expanded significantly. Based on the bibliographic database of the European Project on Ocean Acidification (EPOCA), the number of OA publications averaged about 10–20 per year from 1990 to 2005, and then increased sharply to about 270 publications per year by 2011 (Laffoley and Baxter, 2012). Much of this increase was from studies on biological responses to OA (Nisumaa et al., 2010, 2012). For example, publications of this type accounted for over 80 % of all OA papers in 2011 (Laffoley and Baxter, 2012). With the rapid growth in research and the parallel rise in publications, the need for a comprehensive OA metadata template to facilitate

archiving and access to these data has been recognized by the international ocean acidification data management community (Hansson et al., 2014).

Metadata is structured information that describes an information resource (e.g., an oceanographic data set), enabling its discovery and access (Guenther and Radebaugh, 2004). For OA studies, a metadata record documents such information as what was measured; by whom; when (temporal coverage); where (geographic coverage); how was it sampled and analyzed; with what instruments and following what protocol; and finally what were its units of measure and quality of the data (Pesant et al., 2010).

Metadata is critical to data discovery. It enables data sets to be found through relevant criteria, and provides information about locations of the data sets. Metadata also helps to document information about the data sets, so that they can be understood and utilized beyond the original use. Metadata plays an extremely important role in supporting archiving and preservation of data, facilitating interoperability, and synthesizing legacy data. It serves as the key to ensuring that a data set will be accessible by future researchers (Guenther and Radebaugh, 2004).

While metadata templates for chemical OA data have been available for a long time, the lack of a template for biological response OA data is a significant hindrance to the effective management of these data. Establishing a metadata template is thus fundamental for research into biological responses to OA. We present a metadata template developed in collaboration with many OA researchers. It is applicable to a broad spectrum of OA data sets, including those from studies of biological responses to OA.

2 Methods

2.1 Requirements for the OA metadata template

The envisioned OA metadata template responds to the needs expressed by the community to meet three requirements: (1) to enable data discovery, (2) to document information about OA data sets in a consistent manner, and (3) to be broadly applicable to many types of OA studies. These three requirements served as the guiding principles in the development of the OA metadata template.

2.1.1 Data discovery

One of the main functions of metadata is to enable data discovery. During the OA metadata template development, data discovery was emphasized when decisions were made on the selection of elements and their organization.

2.1.2 Documenting data

Another important role of the metadata template is to document information about a data set. The value of a data set

increases significantly if it comes with the documentation needed to understand and use the data set. If such information could be collected, stored, discovered, and accessed in a consistent manner, the data set would be made more available towards improved assessments of marine ecosystem vulnerability and better OA forecasting capabilities.

2.1.3 Broad applicability

The template targets a broad spectrum of OA data sets. Ocean acidification covers a wide range of oceanographic subject areas, including chemical observations, 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.

2.2 Development process

The OA metadata template development involved two steps:

1. *content standard development* to choose the metadata elements that should be included in the template; and
2. *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).

2.2.1 Content standard development

Developing the content standard involved selecting the metadata elements and establishing their hierarchical relationships. This process involved five steps:

1. studying the experiment setup and analyzing available data sets for biological response OA studies;
2. building on top of existing chemical OA metadata templates;
3. Obtaining feedback from biological OA experts;
4. testing the template by using sample information from publications on biological response OA studies; and
5. finalizing the template and adding definitions for each metadata element.

2.2.2 Format standard development

Numerous metadata format standards exist for documenting geospatial metadata. One of the most complete and widely applicable standards is the International Organization for Standardization (ISO) 191** series, which is designed for geospatial metadata associated with positioning information on the surface of the globe (International Organization for Standardization, 2009). The ISO standards have been

adopted throughout the international environmental community and were officially endorsed as US standards by the Federal Geographic Data Committee (FGDC) in September 2010.

Metadata elements from the content standard development were mapped to their corresponding fields in the ISO 19115-2 standard (International Organization for Standardization, 2009; Mize et al., 2011) to generate the ocean acidification ISO metadata template. The ISO 19115-2 standard was chosen to take advantage of such sections as “MI:Acquisition” which allows data providers to capture information about ships and other data collection platforms in a structured machine readable format. All of the fields in the original ISO 19115 standard are captured within ISO 19115-2, so there is no potential loss of content in choosing to use this version of the standards.

3 OA metadata template

The OA metadata template (Ocean Acidification Data Stewardship Team, 2014) consists of three files:

1. a *submission form* (in Microsoft® Excel spreadsheet) that can be used to prepare the metadata,
2. an *instruction file* (in Adobe® portable document format) that defines the fields and lists their hierarchical relationships, and
3. an *XML file* (encoded with the ISO 19115-2 standard) that can be fed into data search portals for OA data discovery.

In the following sections, the main structure of the metadata template, its elements, and their hierarchical relationships will be described.

3.1 Variable metadata sections

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). Hereafter, we will use the word Variable(s), although Parameter(s) is an acceptable synonym for this discussion. Table 1 lists some commonly used variables in OA research. Variables are treated as the focal point of the entire metadata template because we expect them to be the single most important metadata elements that would be used as search terms to locate a data set. Furthermore, all types of OA research generate some kind of variables, regardless of their sampling scheme, experiment setup, or model inputs. Therefore, the treatment of variables as the focal point of the template allows the template to apply to many types of OA data.

3.1.1 Controlled vocabularies for variables

In order to meet the data discovery goal, it is important to maintain controlled vocabularies for the variables (Table 2). As an example, the variable – dissolved inorganic carbon – is also commonly referred to as total carbon dioxide. If no controlled vocabulary is used, data consumers would have to search the database using all possible variations of the term, in order to locate every available data set. Controlled vocabularies, however, would allow data consumers to locate the data sets by using the corresponding term in the controlled vocabulary.

Ideally, controlled vocabularies should be standardized across all OA data centers. The Ocean Acidification International Coordination Centre (OA-ICC) has been leading the efforts in developing controlled vocabularies for OA data documentation. SeaDataNet, the “Pan-European infrastructure for ocean and marine data management”, also provides controlled vocabularies for many kinds of broader oceanographic services.

3.1.2 Key child elements of a variable

The child metadata elements of a variable are organized around the variable itself to form a “variable metadata section” (Table 2). The following fields, 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 observational context. It could be generic terms that describe how a variable is collected. For example, for chemical OA studies, the observation type could be “Surface underway”, “Time series”, or “Profile” (Table 3). For physiological response OA studies, such terms as “Laboratory experiment”, “Pelagic mesocosm”, “Benthic mesocosm”, or “Natural perturbation site study” could be used (Pesant et al., 2010).

A metadata element called “In-situ/manipulation/response” is also added as a child element of a variable. In physiological response OA studies, variables could fall into several categories. For example, carbon-related variables, e.g., pH, partial pressure of carbon dioxide ($p\text{CO}_2$), dissolved inorganic carbon (DIC), and total alkalinity (TA), are often manipulated to simulate future OA conditions, while other variables, e.g., calcification rate, growth rate, and larval survival rate, are monitored to understand the responses of the organisms to OA. The former group of variables can be categorized as “Manipulation condition variables”, and the latter can be labeled as “Response variables”. To enable this metadata element to be applicable to other types of OA studies, we expanded its scope to cover in-situ observed variables, and call it “In-situ/manipulation/response”.

Table 1. Some commonly used ocean acidification variables, their definitions, and recommended abbreviations.

Variable name [abbreviation or formula]	Definition
Water temperature [Temp]	In situ temperature of the water
Salinity [Salinity] [CTDSAL]	Salinity is the salt content of a water sample or body of water. The measure of salt content of a water sample follows the United Nations Educational Scientific and Cultural Organization (UNESCO) standards known as the Practical Salinity Scale (PSS) as the conductivity ratio of a seawater sample to a standard KCl solution. PSS is a ratio and has no units. Salinity measured from discrete bottle sampling Salinity recorded by the conductivity, temperature, depth (CTD) sensor
Dissolved inorganic carbon [DIC]	Dissolved inorganic carbon (DIC) is the sum of the concentrations of all inorganic carbon species in the ocean: bicarbonate ion (HCO_3^-), carbonate ion (CO_3^{2-}), and un-ionized dissolved carbon dioxide (CO_2^*). $\text{DIC} = [\text{HCO}_3^-] + [\text{CO}_3^{2-}] + [\text{CO}_2^*]$. About 90 % of DIC is present as bicarbonate ion, the proportion of carbonate ion is about a factor of 10 less ($\sim 10\%$), and that of un-ionized carbon dioxide is another factor of 10 less ($< 1\%$). DIC is often inaccurately called total CO_2 , which also includes inorganic CO_2 species bound to organic molecules.
Total alkalinity [TA, Alk, or A_T]	Total alkalinity of seawater is the number of moles of hydrogen ions equivalent to the excess of proton acceptors (bases formed from weak acids with a dissociation constant $K \leq 10^{-4.5}$ at 25 °C and zero ionic strength) over proton donors (acids with $K > 10^{-4.5}$) in 1 kg of sample. $\text{TA} = ([\text{HCO}_3^-] + 2[\text{CO}_3^{2-}] + [\text{B}(\text{OH})_4^-] + [\text{OH}^-] + [\text{HPO}_4^{2-}] + 2[\text{PO}_4^{3-}] + \dots) - ([\text{H}^+] + [\text{HSO}_4^-] + \dots)$. In the oceans, most of the negatively charged ions present are bicarbonate (HCO_3^-) and carbonate (CO_3^{2-}).
pH [pH]	pH is defined as the negative decimal logarithm of the activity of the hydrogen ion in a solution and is a measure of acidity. $\text{pH} = -\log_{10}[\text{H}^+]$. Commonly used pH scales include the free scale, total scale, seawater scale, and National Bureau of Standards (NBS) scale. The free scale includes only the effect of free hydrogen ions. The total scale includes the effect of both free hydrogen ions and hydrogen sulfate ions. The seawater pH scale includes the effect of free hydrogen ions, hydrogen sulfate ions, and fluoride ions. The International Union of Pure and Applied Chemistry (IUPAC) defines a series of buffer solutions across a range of pH values (often denoted with NBS designation). These solutions have a relatively low ionic strength (~ 0.1) compared to that of seawater (~ 0.7), and, as a consequence, are not recommended for use in characterizing the pH of seawater since the ionic strength differences cause changes in electrode potential.
Partial pressure (or fugacity) of carbon dioxide in water [$p\text{CO}_2$ (or $f\text{CO}_2$) water]	The partial pressure of an ideal gas in a mixture is equal to the pressure it would exert if it occupied the same volume alone at the same temperature. The fugacity (f) of a real gas is an effective pressure which replaces the true mechanical pressure in accurate chemical equilibrium calculations. It is equal to the pressure of an ideal gas that has the same chemical potential as the real gas.
Mole ratio of carbon dioxide in the atmosphere [$x\text{CO}_2\text{-atm}$]	The concentration of carbon dioxide in the atmosphere is often recorded as the mole ratio of carbon dioxide molecules and that of all molecules in the atmosphere. Its unit is parts per million (ppm, or ppmv), unlike the units of $p\text{CO}_2$ (or $f\text{CO}_2$) water: micro atmosphere (or μatm).
Dissolved oxygen [Oxygen] [CTDOXY]	The concentration of oxygen dissolved in water Dissolved oxygen measured by chemical analysis (e.g., Winkler titration) Dissolved oxygen measured by optical probes on a CTD rosette
Chlorophyll [CHL]	Chlorophyll is a green pigment found in cyanobacteria and the chloroplasts of algae and plants. All oxygenic photosynthetic organisms use chlorophyll a, but differ in accessory pigments like chlorophyll b.
Chromophoric (or colored) dissolved organic matter [CDOM]	Chromophoric or colored dissolved organic matter is the optically measurable component of the dissolved organic matter in water. It occurs naturally in aquatic environments primarily as a result of tannins released from decaying detritus.

Table 1. Continued.

Variable name [abbrev. or formula]	Definition
Photosynthetically active radiation [PAR]	Photosynthetically active radiation is a measurement of the spectral range (wave band) of solar radiation from 400 to 700 nm that photosynthetic organisms are able to use in the process of photosynthesis.
Nitrate [Nitrate]	In seawater, inorganic nitrogen is found in several different forms, including ammonia/ammonium ($\text{NH}_3/\text{NH}_4^+$), nitrates (NO_3^-), and nitrites (NO_2^-). The nitrate ion is the base of nitric acid (HNO_3), with a molecular formula of NO_3^- . It is the principal form of fixed dissolved inorganic nitrogen assimilated by organisms.
Nitrite [Nitrite]	The nitrite ion (NO_2^-) is a symmetric anion with equal N–O bond lengths. Upon protonation, the unstable weak acid nitrous acid (HNO_2) is produced.
Nitrite + nitrate [Nitrite + nitrate]	Nitrate is often quantitatively reduced to nitrite, before the reduced nitrite as well as the previously existing nitrite in seawater can be determined colorimetrically. The sum of the nitrite (NO_2^-) and nitrate (NO_3^-) concentrations is often reported when this analytical method is used.
Ammonia/ammonium [Ammonia/ammonium]	Ammonia and ammonium are acid/base pairs. Ammonia is a compound of nitrogen and hydrogen with the formula NH_3 . The ammonium cation is a positively charged polyatomic ion with the chemical formula NH_4^+ . It is formed by the protonation of ammonia (NH_3).
Phosphate [Phosphate]	Phosphoric acid is a mineral (inorganic) acid, and the chemical formula is H_3PO_4 . The conjugate base of phosphoric acid is the dihydrogen phosphate ion (H_2PO_4^-) that in turn has a conjugate base of hydrogen phosphate (HPO_4^{2-}), which has a conjugate base of phosphate (PO_4^{3-}). Phosphate, a salt of phosphoric acid, is considered the most important phosphorus species that is immediately biologically available in seawater. Phosphate exists predominantly in the form of HPO_4^{2-} in seawater.
Silicate [Silicate]	Silicon exists in seawater usually as silica (SiO_2) or silicates (SiO_4^{4-} or SiO_3^{2-}). Silicate is commonly used to refer to the sum of silica and silicate. Dissolved silicate concentrations in seawater range from $< 1 \mu\text{mol kg}^{-1}$ in surface waters to $\sim 180 \mu\text{mol kg}^{-1}$ in the deep North Pacific.
Chlorofluorocarbons [CFC]	Chlorofluorocarbons are gases that are synthetic halogenated methanes. They were introduced as industrial coolants in the 1930s and afterward. In oceanography, they are used as tracers of ocean circulation.
CFC-11	The chemical structure of CFC-11 is CCl_3F .
CFC-12	The chemical structure of CFC-12 is CCl_2F_2 .
CFC-113	The chemical structure of CFC-113 is $\text{CCl}_2\text{FCClF}_2$.
Delta carbon-13 [DELC13]	Delta carbon-13 ($\delta^{13}\text{C}$) is a measure of the ratio of the carbon isotopes $^{13}\text{C} : ^{12}\text{C}$ (carbon-13 : carbon-12) in the sample to that in the reference standard, reported in parts per thousand (per mil, ‰).
Delta carbon-14 [DELC14]	Delta carbon-14 ($\Delta^{14}\text{C}$) is a measure of the ratio of carbon isotopes $^{14}\text{C} : ^{12}\text{C}$ (carbon-14 : carbon-12) in the sample to that in the reference standard, reported in parts per thousand (per mil, ‰). $\Delta^{14}\text{C}$ represents the “normalized” value of $\delta^{14}\text{C}$. Normalized means that the effect of fractionation is removed.
Delta oxygen-18 [DELO18]	Delta oxygen-18 ($\delta^{18}\text{O}$) is a measure of the ratio of oxygen isotopes $^{18}\text{O} : ^{16}\text{O}$ (oxygen-18 : oxygen-16) in the sample to that in the reference standard.
Delta nitrogen-15 [DELN15]	Delta nitrogen-15 ($\delta^{15}\text{N}$) is a measure of the ratio of nitrogen isotopes $^{15}\text{N} : ^{14}\text{N}$ (nitrogen-15 : nitrogen-14) in the sample to that in the reference standard.

Table 2. 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
	Controlled vocabulary name
	Observation type
	In-situ/manipulation/response variable
	Manipulation method
	Variable unit
	Measured or calculated
	Calculation method and parameters
	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)
	Life stage of the biological subject
	Researcher name
	Researcher institution

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 species on which the larval survival rate was studied. An element, called “Biological subject”, is where users identify an organism or a biological community, to which the observation applies.

3.1.3 Additional child elements of a variable

The four metadata elements discussed above – variable name, observation type, in-situ/manipulation/response, and biological subject – form the skeleton structure of any “variable metadata section”. They are also the main discovery metadata elements that would be used to locate data in a data search portal. However, additional metadata elements are needed to make the variable independently understandable.

“Variable abbreviation” documents the abbreviation or formula of a variable in the data files. “Full variable name” spells out the detailed descriptive name of the variable. For manipulation condition variables, their manipulation methods, e.g., bubbling carbon dioxide, adding acid or base to the solution, can be recorded in a field called “Manipulation method”. “Units” should be reported in accordance with the National Institute of Standards and Technology (NIST) International System of Units (SI). In addition, whether a variable is “Measured or calculated”, and its “Calculation method and

parameters” are also important pieces of information to document (Table 2).

Instrumentation is split into two categories: “Sampling instruments” and “Analyzing instruments”. A common mistake in data management practices is that sampling and analyzing instruments are used interchangeably. For example, if a researcher measured dissolved oxygen (DO) using an oxygen sensor attached to a conductivity, temperature, depth (CTD) rosette, some people may document the instrument of the study as CTD rosette, but others may think the instrument is the oxygen sensor, even though both should be recorded: CTD rosette as the sampling instrument and oxygen sensor as the analyzing instrument. Instruments that are used to collect water samples or deploy sensors are here defined as sampling instruments. Examples of sampling instruments include a CTD rosette, a Niskin bottle, and a flow-through pump onboard a research vessel. The term Analyzing instruments, however, refers to instruments that are used to analyze water samples collected with the sampling instruments, or sensors that are mounted on the sampling instruments to measure some variables of the water. For example, the analyzing instrument for a pH measurement could be a glass electrode coupled with a pH meter, or a spectrophotometer. In addition, we also created a free text field called “Detailed sampling and analyzing information” to allow users to capture additional details of their sampling and analyzing procedures beyond what instruments are used.

Several elements that elaborate on the sample size and data quality are also included in the template. “Uncertainty” is an open-text field that allows users to document information about the data quality of the variable. Input to this field could be the standard deviation of the measurements (e.g., 1 %, $2 \mu\text{mol kg}^{-1}$), instrument error logs, or other information related to the quality control of the variable. The data quality flag conventions/rules of the variables can be stored in the field “Data quality flag description”. Input to this field includes the name of the data quality flag convention or descriptions of the data quality flag rules, e.g., “2-good quality data”, “3-questionable data”, and “4-bad data”.

In addition to “Biological subject”, a field called “Species identification code” is added to document the standard species IDs, if such information is available (Table 2). Using the reference databases from the Integrated Taxonomic Information System (<http://www.itis.gov/>) or World Register of Marine Species (<http://marinespecies.org/>) is recommended, although use of any other reference databases is allowed. The name of the reference database can also be recorded in this metadata field. The field “Life stage of the biological subject” documents development stages of the species. Input to this field can be eggs, larvae, juveniles, adults, etc. The use of the British Oceanographic Data Center (BODC) parameter semantic model biological entity development stage terms (S11) is recommended.

“Method reference” is reserved to hold the bibliographic citation information of the method used to measure the

Table 3. 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 $p\text{CO}_2$ 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 raise the 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)
Benthic 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). “Benthic” zone is the ecological region at the lowest level of a body of water such as an ocean or a lake.	(none)
Natural perturbation site study	The natural perturbation site study is based on looking directly at how organisms and communities and ecosystems react to high/low pH and carbonate saturation state in the real world, replete with all its biodiversity, ecosystem interactions, and adaptation to the ambient chemistry.	(none)
Model output	Generation of data from numerical models	(none)

variable. Most OA studies involve collaboration from multiple scientists. Therefore, it is important to document the researchers’ information for each variable, to give them credit for their sampling and analyzing efforts and to allow trace-

ability for future questions. For the sake of brevity, only researchers’ names and their affiliated institutions are recorded.

3.2 Investigators

“Investigators” information is needed to credit researchers for their overall data collection and analysis efforts. In addition to the basic address and contact information, we recommend the use of personal identifiers (e.g., ORCID, ResearcherID) to unambiguously define the investigator, and recommend using a controlled vocabulary for organizations as well.

3.3 Temporal and spatial coverage

Temporal and spatial information provide important data constraints. The use of ISO-8601 date (YYYY-MM-DD, e.g., 1997-07-16) and date plus hours, minutes, and seconds (YYYY-MM-DDThh:mm:ssTZD, e.g., 1997-07-16T19:20:30+01:00, TZD is Time Zone Designator) formats are recommended. Bounding box coordinates refer to the rectangular box whose edges are defined by the northernmost and southernmost latitudes, and the westernmost and easternmost longitudes. “Geographic names” are the names of the seas or water bodies where the sampling takes place. It could be as broad as an ocean basin, such as North Pacific Ocean, or as specific as a local river or bay. For biological studies, the bounding box coordinates and geographic names are used to document the location of the water collection. An additional field called “Location of organism collection” is used to document where the organisms were collected.

3.4 Platforms and sampling IDs

“Platforms” often refer to the research vessels that carry out the research. However, platforms could be something other than a ship (e.g., glider, Argo, satellite) or something that is fixed (e.g., moored buoys, towers). “Expedition code (EX-POCODE)” consists of the four-digit International Council for the Exploration of the Sea (ICES) platform code and the sailing date in the YYYYMMDD format. For a list of ICES platform codes, please refer to <http://vocab.ices.dk>. “Section ID” is the identification number for a research cruise section or leg. It was commonly used during the WOCE studies, which often had many repeating cruises on a single section, e.g., A16N. “Cruise ID” is the particular ship cruise number (e.g., MT901), or other alias for the cruise. For example, a cruise ID (e.g., A16N_2013) could consist of a section ID (e.g., A16N) and the sampling year (e.g., 2013).

3.5 Citation, references, and supplementary information

“Citation” refers to bibliographic information about how the data set should be cited. When working with a data center to publish their data, data producers may only need to provide the author list to complete the citation as the other portions of the citation are typically the responsibility of the data publisher. When compiling an author list, we recommend using the format of Lastname1, Firstname1 Middle-

name1; Lastname2, Firstname2 Middlename2; ... For data centers, we recommend the use of the styles described in the FORCE11 Joint Declaration of Data Citation Principles (<https://www.force11.org/datacitation>) to organize your data citations.

“References” are bibliographic citations of publications, e.g., papers, cruise reports, that describe the data set. Researchers often submit their data to an archive after their work has been published. It is important to share related publications to help future users better understand the data set. The “Supplementary information” field is reserved for any information critical to understanding the data set that does not fit into any other existing fields.

4 Conclusions

The creation of a common metadata template to manage biological response OA data sets is a major effort by the OA research and data management community. We described a metadata template that applies to many types of OA data sets, including chemical OA data sets and those describing the biological responses to OA. In addition to serving OA data management efforts effectively, the template can be used by the OA research community for documenting their OA data sets, sharing data sets among researchers, and submitting data sets to data centers. The metadata template files are stored at the National Oceanic and Atmospheric Administration institutional repository with a digital object identifier (DOI) of 10.7289/V5C24TCK. The metadata development approach documented here can benefit other scientific data management programs in terms of metadata template development.

Acknowledgements. This work was supported by the Ocean Acidification Program (OAP) of the National Oceanic and Atmospheric Administration (NOAA). Discussions with Hernan E. Garcia (National Centers for Environmental Information) benefitted the metadata template development. We thank Philip Goldstein (University of Colorado) for his contribution to the metadata template development and his comments on an earlier draft of the paper. Alex Kozyr (Carbon Dioxide Information Analysis Center) provided help with the documentation of chemical OA data sets. We are grateful to Andrew Dickson of Scripps Institution of Oceanography and Fiz Pérez of the Spanish National Research Council for their insightful comments on the template. Jacqueline Mize (National Coastal Data Development Center) and Sheri Philips (National Centers for Environmental Information) provided tremendous help on the use of the ISO 19115-2 format and the creation of the ISO version of the template. We thank the associate editor, Robert Key; two reviewers, Cynthia Chandler and Anton Velo; and two readers, J.-P. Gattuso and Yan Yang, for their excellent comments that helped to improve both the template and the paper. We thank Linda Jenkins (National Coastal Data Development Center) for her technical editing. We also thank Chris Chambers, Thomas Hurst, Chris Long, Annette DesRochers, Molly Timmers, Nina Bednaršek, Donald Christopher Melrose,

Derek Manzello, Renee Carlton, Jessica Morgan (NOAA), and David Kline (Scripps Institution of Oceanography) for their comments on the template. Rob Ragsdale (US Integrated Ocean Observing System), Emilio Mayoga (University of Washington), and Sara Haines (University of North Carolina) contributed to the development of definitions for some of the variables. We are indebted to Dean Perry and Dylan Redman (NOAA Northeast Fisheries Science Center) for allowing us to use their metadata records as a real world example of our metadata submission form. We thank Mark Fornwall (United States Geological Survey) for his comments to improve the paper.

Edited by: R. Key

References

- Caldeira, K. and Wickett, E.: Anthropogenic carbon and ocean pH, *Nature*, 425, 365, 2003.
- Feely, R. A., Sabine, C. L., Lee, K., Berelson, W., Kleypas, J., Fabry, V. J., and Millero, F. J.: Impact of anthropogenic CO₂ on the CaCO₃ system in the oceans, *Science*, 305, 362–366, 2004.
- Guenther, R. and Radebaugh, J.: *Understanding Metadata*, National Information Standards Organization (NISO) Press, Bethesda, MD, ISBN 1-880124-62-9, 2004.
- Hansson, L., Appletans, W., and Gattuso, J.-P.: Ocean acidification data management – workshop initiates first steps towards international collaboration, *Eos*, 95, 421–422, 2014.
- Hönisch, B., Hemming, N. G., Archer, D., Siddall, M., and McManus, F.: Atmospheric carbon dioxide concentration across the mid-Pleistocene transition, *Science*, 324, 1551–1554, 2009.
- International Organization for Standardization (ISO): *Geographic information – metadata – Part 2, Extensions for imagery and gridded data*, Geneva, Switzerland, 2009.
- Laffoley, D. d’A. and Baxter, J. M. (Eds.): *Ocean acidification: the knowledge base 2012. Updating what we know about ocean acidification and key global challenges*. European Project on Ocean Acidification (EPOCA), UK Ocean Acidification Research Programme, (UKOA), Biological Impacts of Ocean Acidification (BIOACID) and Mediterranean Sea Acidification in a Changing Climate (MedSeA), 8 pp., 2012.
- Le Quéré, C., Peters, G. P., Andres, R. J., Andrew, R. M., Boden, T. A., Ciais, P., Friedlingstein, P., Houghton, R. A., Marland, G., Moriarty, R., Sitch, S., Tans, P., Armeth, A., Arvanitis, A., Bakker, D. C. E., Bopp, L., Canadell, J. G., Chini, L. P., Doney, S. C., Harper, A., Harris, I., House, J. I., Jain, A. K., Jones, S. D., Kato, E., Keeling, R. F., Klein Goldewijk, K., Körtzinger, A., Koven, C., Lefèvre, N., Maignan, F., Omar, A., Ono, T., Park, G.-H., Pfeil, B., Poulter, B., Raupach, M. R., Regnier, P., Rödenbeck, C., Saito, S., Schwinger, J., Segschneider, J., Stocker, B. D., Takahashi, T., Tilbrook, B., van Heuven, S., Viovy, N., Wankhoff, R., Wiltshire, A., and Zaehle, S.: Global carbon budget 2013, *Earth Syst. Sci. Data*, 6, 235–263, doi:10.5194/essd-6-235-2014, 2014.
- Mize, J., Milan, A., Ball, A., Fergusson, E., Martinolich, K., Jenkins, K., Wayne, L., Schweitzer, P., Herndon, P., O’Connor, S., Mesick, S., Habermann, T., and Hutchison, V.: *Guide to Implementing ISO 19115-2:2009(E), the North American Profile (NAP), and ISO 19110 Feature Catalogue*, National Coastal Data Development Center, Stennis Space Center, Mississippi, 2011.
- Nisumaa, A.-M., Pesant, S., Bellerby, R. G. J., Delille, B., Middelburg, J. J., Orr, J. C., Riebesell, U., Tyrrell, T., Wolf-Gladrow, D., and Gattuso, J.-P.: EPOCA/EUR-OCEANS data compilation on the biological and biogeochemical responses to ocean acidification, *Earth Syst. Sci. Data*, 2, 167–175, doi:10.5194/essd-2-167-2010, 2010.
- Nisumaa, A.-M., Schlitzer, R., Hansson, L., and Gattuso, J.-P.: EPOCA data management activities: a summary, *IMBER Newsletter* 23, 2012.
- Ocean Acidification Data Stewardship Team: *Ocean acidification metadata template*, National Oceanic and Atmospheric Administration institutional repository, Accession Number: ocn881471371, doi:10.7289/V5C24TCK, 2014.
- Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R. M., Lindsay, K., Maier-Reimer, E., Matear, R. J., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G.-K., Rodgers, K. B., Sabine, C. L., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Totterdell, I. J., Weirig, M.-F., Yamanaka, Y., and Yool, A.: Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms, *Nature*, 437, 681–686, 2005.
- Pesant, S., Alan Hook, L., Lowry, R., Moncoiffé, G., Nisumaa, A.-M., and Pfeil, B.: *Safeguarding and sharing ocean acidification data*, in: *Guide to best practices for ocean acidification research and data reporting*, edited by: Riebesell, U., Fabry, V. J., Hansson, L., and Gattuso, J. P., Publications Office of the European Union, Luxembourg, 260 pp., 2010.
- Riebesell, U., Lee, K., and Nejstgaard, J. C.: *Pelagic mesocosms*, in: *Guide to best practices for ocean acidification research and data reporting*, edited by: Riebesell, U., Fabry, V. J., Hansson, L., and Gattuso, J. P., Publications Office of the European Union, Luxembourg, 260 pp., 2010.