Earth Syst. Sci. Data Discuss., 8, 1–21, 2015
www.earth-syst-sci-data-discuss.net/8/1/2015/
doi:10.5194/essdd-8-1-2015

Open Access Earth System Science Data Discussions

A metadata template for ocean acidification data

L.-Q. Jiang[1], S. A. O'Connor[2], K. M. Arzayus[3], A. Kozyr[4], and A. R. Parsons[3]

[1]Cooperative Institute for Climate and Satellites-Maryland, Earth System Science Interdisciplinary Center, University of Maryland, 5825 University Research Ct., College Park, Maryland 20740, USA
[2]ADNET Systems, Inc., 6720B Rockledge Dr., Bethesda, Maryland 20817, USA
[3]National Oceanographic Data Center, National Oceanic and Atmospheric Administration, 1315 East–West Highway, Silver Spring, Maryland 20910, USA
[4]Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, 1 Bethel Valley Rd., Oak Ridge, Tennessee 37831, USA

Received: 18 November 2014 – Accepted: 8 December 2014 – Published: 13 January 2015

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

Published by Copernicus Publications.

Abstract

This paper defines the 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 responses 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 paper, we present a metadata template that can be applied to a broad spectrum of OA studies, including those studying the biological responses 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 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. 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 (CO_2) into the atmosphere, causing the atmospheric CO_2 to increase by over 40 % compared to the highest CO_2 levels in the last 800 000 years (Hönisch et al., 2009). However, only about half of the CO_2 emissions remain in the

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

ESSDD

8, 1–21, 2015

OA metadata template

L.-Q. Jiang et al.

Title Page
Abstract Instruments
Data Provenance & Structure
Tables Figures
◀▶ ◀ ▶
Back Close
Full Screen / Esc

Printer-friendly Version

Interactive Discussion

atmosphere, the remaining portion is taken up by the ocean and terrestrial systems. Based on the most recent emission data from 2004 to 2013, the global oceans take up about 26 % of all the CO_2 emissions from anthropogenic activities (Le Quéré et al., 2014). The oceans' uptake of atmospheric CO_2 is causing decreasing pH, carbonate ions (CO_3^{2-}) concentration, and carbonate saturation state, a process commonly referred to as "ocean acidification (OA)" (Caldeira and Wickett, 2003; Feely et al., 2004; Orr et al., 2005).

Over the past 10 years, the number of scientific studies undertaken to understand OA has increased significantly. 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 responses 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 responses 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.

Metadata is structured information that describes an information resource (e.g., an oceanographic data set), enabling its discovery and availability for use (Guenther and Radebaugh, 2004). For OA studies, metadata 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 (Pesant et al., 2010).

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. It also 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 plays

ESSDD
8, 1–21, 2015

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 and Radebaugh, 2004).

While metadata templates for water chemistry OA data have been available for a long time, the lack of a metadata template for biological response OA data has become a significant hindrance for the effective management of this type of data. Establishing a metadata template for them has emerged as a fundamental element for furthering research into the study of biological responses of organisms to OA. In this paper, we present a metadata template that was developed in collaboration with many OA researchers. It is applicable to a broad spectrum of OA data sets, including those that describe the biological responses of organisms to OA.

2 Methods

2.1 Requirements for the OA metadata template

The envisioned OA metadata template responds to expressed needs of scientists and data managers to meet three requirements: (1) to enable data discovery, (2) to document information about OA data sets in consistent and standard ways, and (3) to be broadly applicable to many types of OA studies with the goal of creating a template that can be adopted by the international community. 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 goals of any metadata template is to enable data discovery. The developed OA metadata template will be formatted according to the International Organization for Standardization (ISO) metadata standards, and then fed into data discovery

ESSDD
8, 1–21, 2015

portals to ensure OA data sets are easily discoverable. During the OA metadata template development, data discovery was always emphasized when decisions were made on the elements to be included in the template, and how these elements were to be organized.

2.1.2 Documenting data

Another important role of the metadata template is to document as much useful information of a data set as consistently as possible. The values of a data set increase significantly if it comes with a metadata record containing all the information needed to understand and use the data set. If such metadata information can be collected, stored, discovered and accessed in consistent and standard ways, 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 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. In addition, we apply international standards wherever possible to benefit both the international research community and data repositories of the future.

2.2 Development process

The development of the OA metadata template involved two steps:

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

ESSDD
8, 1–21, 2015

L.-Q. Jiang et al.

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 that should be included in the template and then building 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 the existing metadata templates for water chemistry OA data;
3. working with biological OA experts from the research community for their feedback;
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

A number of metadata format standards exist for documenting environmental geospatial metadata. One of the most complete and widely applicable standards is the International Organization for Standardization (ISO)'s 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 placed to their corresponding XML fields according to the ISO 19115-2 standards (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 will allow data producers 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 also captured within ISO 19115-2, so there is no potential loss of contents in choosing to use the later extension to the originally issued ISO standard.

3 OA metadata template

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

1. an *instruction file* (in Adobe® portable document format) that lists the hierarchical relationships of all the metadata elements and their definitions;
2. a *submission form* (in Microsoft® Excel spreadsheet) that can be used by data providers to prepare their metadata; and
3. an *XML file* (encoded with the ISO 19115-2 standards) 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.

ESSDD
8, 1–21, 2015
OA metadata template
L.-Q. Jiang et al.

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. Variables are treated as the focal point of the entire metadata template, because we expect them to be the single, most important metadata element that would be used as search terms to locate an OA data set. 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 could potentially allow the template to fit many types of OA studies. Table 1 lists some commonly used variables in OA studies.

3.1.1 Key child elements of a variable

The child metadata elements of a variable were organized around the variable itself to form a "variable metadata section" (Table 2). 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 will establish 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 water chemistry 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.

A metadata element called "in-situ/manipulation/response" was also added as a child element of a variable. In physiological response OA studies, variables could be classified into several categories. For example, carbon-related variables, e.g., pH, partial pressure of carbon dioxide (pCO$_2$), dissolved inorganic carbon (DIC), and total alkalin-

ESSDD
8, 1–21, 2015

ity (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 ocean acidification. 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".

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. The element, called "biological subject", is where users identify an organism or a biological community, to which the observation of the variable applies.

3.1.2 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 by data search portals to conduct their searches. However, that information alone cannot document an OA variable with full details. More metadata elements are needed to complete the "variable metadata section".

"Variable abbreviation" was added to document the abbreviation or formula of a variable in the data file, so that future users will always know what the abbreviations in the data files mean. Its "Climate and Forecast standard name" is also recorded, if it is available. In addition, the "unit" of the studied variable, and whether a variable is "measured or calculated" 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

ESSDD
8, 1–21, 2015
OA metadata template
L.-Q. Jiang et al.

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, but others may think the instrument is the oxygen sensor, even though both should be recorded: CTD 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, 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 instrument to measure some variables of the water. For example, the analyzing instrument for a pH measurement could be either a glass electrode coupled with a pH meter or a spectrophotometer. The latter can often measure pH much more accurately than the former. If sampling instruments shed light on the sampling scheme, analyzing instruments give users information about the data quality.

Sampling and analyzing instruments contribute to the understanding of the sampling and analyzing scheme as well as the accuracy of the measurement. However, there are many more sampling and analyzing details than the instrumentation information alone. For that reason, we also created a free text field called "detailed sampling and analyzing information" to allow users to document any other details of their sampling and analyzing procedures.

Several elements that elaborate on the sample size and data quality are also included in the template. Such information as how many replicate samples are taken and analyzed at a certain time and location could be very informative about the statistical certainty and accuracy of the measurement. The element "field replicate information" was added to capture information about the number of replicates. "Uncertainty" is an open field that allows users to document information about the data quality of the variable. Input to this field could be the SD of the measurements (e.g., 1 %, 2 $\mu mol\,kg^{-1}$), instrument error logs, or other information related to the quality control of the variable.

ESSDD
8, 1–21, 2015

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 a specific data quality flag convention or the contents of the data quality flag rules, e.g., "2 – good quality data, 3 – questionable, 4 – bad data".

In addition to "biological subject", a field called "species identification code" was added to document the standard species IDs, if such information is available. 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 also allowed. The name of the reference database can also be placed in this metadata field.

Different methods could be used to measure a single variable. "Method reference" is reserved to hold the bibliographic citation information of the method used to analyze the samples. Most OA studies involve collaboration from multiple scientists. Chances are that some variables will show as having been measured by Researcher A, others measured by Researcher B, and so on. 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 traceability for future questions. For the sake of brevity, only researchers' names and their affiliated institutions are recorded for each variable.

3.2 Investigators

In addition to the "variable metadata sections" discussed previously, many other metadata elements are included to complete the template. Investigators information is needed to credit researchers for their data collection and analyses efforts. Recent studies indicated that publications whose data sets are also made publicly available receive a higher number of citations than similar studies without publicly available data (Dorch, 2012; Piwowar and Vision, 2013). The research community has been moving toward a direction where researchers receive credit not only by publishing papers, but also by publishing their data. Standard practices are evolving such that each data set would be published with its own citation details and digital object identifier (DOI), thereby creating

ESSDD
8, 1–21, 2015

the potential for data citation reports to be generated that display all the publications that cite the data set, similar to those presented by Thomson Reuters® "Web of Science" for paper publications. Therefore, it is important to document the investigators and their related information in the metadata of each data set.

3.3 Temporal and spatial coverage

Temporal coverage information (i.e., the "start date", "end date") and spatial coverage information (i.e., the "bounding box coordinates" and "geographic names") are important for constraining the sampling dimensions of a data set. The world oceans are a dynamic system that varies significantly in space and time. All oceanographic studies require such dimensional information to accurately interpret what is observed. Start date and end date are defined as the first and last day of the observation, respectively. Bounding box coordinates refer to the rectangular box whose corners 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. The bounding box coordinates and geographic names for laboratory experiment studies have been undergoing rigorous debates recently. The general consensus is that they should be used to document the location of the water collection (S. Meseck, personal communication, 2014).

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) or something that is fixed (e.g., moored buoys, towers). Expedition code (EXPOCODE) consists of the four-digit International Council for the Exploration of the Sea (ICES) ship code and the date of the first day of the cruise in the YYYYMMDD format. The ICES ship code is interchangeable with the National Oceanographic Data Center (NODC)'s platform code most of the time. When discrepancies occur, the community tends to follow the ICES code. The first two characters of the ICES ship code are the country code, and the other two are used to identify an individual vessel. For a list of ICES ship and country 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 World Ocean Circulation Experiment (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, the cruise ID (e.g., A16N_2013) could consist of a Section ID (e.g., A16N), and the sampling year (e.g., 2013).

3.5 Funding agencies and projects

Funding Agency is used to acknowledge the financial support from agencies or organizations. This field may become increasingly important, as it could facilitate the enforcement of White House Executive Order: "Making Open and Machine Readable the New Default for Government Information " (9 May 2013), which requires all federal funding agencies to make data collected from their funded research openly available. Project refers to a collaborative research effort that is carefully planned and designed to achieve a particular research goal. Examples include Climate Variability and Predictability (CLIVAR), Tropical Atmosphere Ocean (TAO), and U.S. Joint Global Ocean Flux Study (US JGOFS).

3.6 Data citation, references, and supplementary information

Data Citation refers to bibliographic information about how the data set should be cited. It is similar to a publication citation. For data submission, a data provider can simply provide their author list in the correct order in the "Data citation" field. We recommend using the format of Lastname1, Firstname1 Middlename1; Lastname2, Firstname2 Middlename2; . . . for the author list. References are bibliographic citations of publications, e.g., papers, cruise reports, etc., that describe the data set. Researchers often

ESSDD
8, 1–21, 2015

submit their data to an archive after their conducted research has been published. Therefore, it is important to encourage them to share related publications wherever possible to help future users better understand the data set. The supplementary information field is reserved for any information critical to understanding the data set, but 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 numerous types of OA data sets, including water chemistry data sets and those describing the biological responses of organisms on 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 template, as well as its submission form, and ISO 19115-2 XML file 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. We gratefully acknowledge the funding support from the National Oceanic and Atmospheric Administration (NOAA)'s Ocean Acidification Program (OAP). Discussions with Hernan E. Garcia (NODC) enhanced 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. Sheri Philips (NODC), and Jacqueline Mize (National Coastal Data Development Center) offered tremendous help on the use of the ISO 19115-2 format and the creation of the ISO version of the template. We are grateful to Linda Jenkins (National Coastal Data Development Center) for her technical editing that improved the paper significantly. We also thank Chris Chambers, Thomas Hurst, Chris Long, Annette DesRochers, Molly Timmers, Nina Bednaršek, Derek Manzello, Renee Carlton, and Jessica

ESSDD
8, 1–21, 2015

Morgan (NOAA), as well as David Kline (Scripps Institution of Oceanography) for their comments on the template. We thank Mark Fornwall (United States Geological Survey) for his comments to improve the paper. Rob Ragsdale (U.S. 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 record as a real world example of our metadata submission form.

References

Barry, J. P., Hall-Spencer, J. M., and Tyrrell, T.: In situ perturbation experiments: natural venting sites, spatial/temporal gradients in ocean pH, manipulative in situ $p(CO_2)$ perturbations, 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.

Caldeira, K. and Wickett, E.: Anthropogenic carbon and ocean pH, Nature, 425, p. 365, doi:10.1038/425365a, 2003.

Dorch, B.: On the citation advantage of linking to data, Hprints, available at: http://hprints.org/hprints-00714715 (last access: 15 November 2014), 2012.

Feely, R. A., Sabine, C. L., Lee, K., Berelson, W., Kleypas, J., Fabry, V. J., and Millero, F. J.: Impact of anthropogenic CO_2 on the $CaCO_3$ system in the oceans, Science, 305, 362–366, 2004.

Guenther, R. and Radebaugh, J.: Understanding Metadata, National Information Standards Organization (NISO) Press, Bethesda, MD, 2004.

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.

Le Quéré, C., Moriarty, R., Andrew, R. M., Peters, G. P., Ciais, P., Friedlingstein, P., Jones, S. D., Sitch, S., Tans, P., Arneth, A., Boden, T. A., Bopp, L., Bozec, Y., Canadell, J. G., Chevallier, F., Cosca, C. E., Harris, I., Hoppema, M., Houghton, R. A., House, J. I., Jain, A.,

ESSDD
8, 1–21, 2015

OA metadata template

L.-Q. Jiang et al.

Johannessen, T., Kato, E., Keeling, R. F., Kitidis, V., Klein Goldewijk, K., Koven, C., Landa, C. S., Landschützer, P., Lenton, A., Lima, I. D., Marland, G., Mathis, J. T., Metzl, N., Nojiri, Y., Olsen, A., Ono, T., Peters, W., Pfeil, B., Poulter, B., Raupach, M. R., Regnier, P., Rödenbeck, C., Saito, S., Salisbury, J. E., Schuster, U., Schwinger, J., Séférian, R., Segschneider, J., Steinhoff, T., Stocker, B. D., Sutton, A. J., Takahashi, T., Tilbrook, B., van der Werf, G. R., Viovy, N., Wang, Y.-P., Wanninkhof, R., Wiltshire, A., and Zeng, N.: Global carbon budget 2014, Earth Syst. Sci. Data Discuss., 7, 521–610, doi:10.5194/essdd-7-521-2014, 2014.

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, in: 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.

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.

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.

Piwowar, H. A. and Vision, T. J.: Data reuse and the open data citation advantage, PeerJ, 1, e175, doi:10.7717/peerj.175, 2013.

ESSDD
8, 1–21, 2015

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.

ESSDD
8, 1–21, 2015

OA metadata template

L.-Q. Jiang et al.

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

Variable name [abbrev. or formula]	Definition
Water temperature [Temp]	In situ temperature of the water.
Salinity [Sal]	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.
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^*). DIC = $[HCO_3^-]+[CO_3^{2-}]+[CO_2^*]$. About 90 % of DIC is present as bicarbonate ion, the proportion of carbonate ion is about a factor of 10 less (~ 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 ion 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. TA = $([HCO_3^-]+2[CO_3^{2-}]+[B(OH)_4^-]+[OH^-]+[HPO_4^{2-}]+2[PO_4^{3-}]+\ldots)-([H^+]+[HSO_4^-]+\ldots)$. 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. pH = $\log_{10}[H^+]$. Commonly used pH scales include: 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 [pCO$_2$ (or fCO$_2$)-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 [xCO$_2$-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 pCO$_2$ (or fCO$_2$)-water: micro atmosphere (or µatm).
Dissolved oxygen [DO]	Concentration of dissolved oxygen (DO) in water can be measured by laboratory chemical analysis (e.g., Winkler titration), electrochemical or optical probes. DO concentration is highly dependent on temperature and salinity.
Oxygen – percent saturation [DO (%)]	Percent saturation is the amount of oxygen in a certain volume of water relative to the total amount of oxygen that the water can hold at that temperature and salinity. It is often calculated as the ratio between measured dissolved oxygen and theoretical saturated oxygen content estimated from temperature and salinity.
Apparent oxygen utilization [AOU]	Apparent oxygen utilization is the difference between the measured dissolved oxygen concentration and its equilibrium saturation concentration in water with the same physical and chemical properties. Such differences typically occur when biological activity acts to change the ambient concentration of oxygen.
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 chlorophylls *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.
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 [NITRAT]	In seawater, inorganic nitrogen is found in several different forms, including ammonia/ammonium (NH_3/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.

ESSDD
8, 1–21, 2015

Table 1. Continued.

Variable name [abbrev. or formula]	Definition
Nitrite [NITRIT]	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 [NITRIT + NITRAT]	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 [PHSPHT]	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 pre-dominantly in the form of HPO_4^{2-} in seawater.
Silicate [SILCAT]	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 $< 1\,\mu mol\,kg^{-1}$ in surface waters to $\sim 180\,\mu mol\,kg^{-1}$ in the deep North Pacific.
Chlorofluorocarbons [CFC] CFC-11 CFC-12 CFC-113	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. The chemical structures of CFC-11 is CCl_3F. The chemical structures of CFC-12 is CCl_2F_2. The chemical structures of CFC-113 is CCl_2FCClF_2.
Delta carbon 13 [DELC13]	Delta carbon-13 ($\delta^{13}C$) is a measure of the ratio of the carbon isotopes $^{13}C:{}^{12}C$ (Carbon-13 : Carbon-12), reported in parts per thousand (per mil, ‰).
Delta carbon 14 [DELC14]	Delta carbon-14 ($\delta^{14}C$) is a measure of the ratio of carbon isotopes $^{14}C:{}^{12}C$ (Carbon-14 : Carbon-12), reported in parts per thousand (per mil, ‰).
Delta oxygen 18 [DELO18]	Delta oxygen 18 ($\delta^{18}O$) is a measure of the ratio of oxygen isotopes $^{18}O:{}^{16}O$ (Oxygen-18 : Oxygen-16).
Delta nitrogen 15 [DELN15]	Delta nitrogen 15 ($\delta^{15}N$) is a measure of the ratio of nitrogen isotopes $^{15}N:{}^{14}N$ (Nitrogen-15 : Nitrogen-14).

ESSDD
8, 1–21, 2015

ESSDD
8, 1–21, 2015

Table 2. Variable metadata section, with child metadata elements organized around the variable/parameter. The elements in bold fonts form the skeleton structure of the variable metadata section.

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

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 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)
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 saturation state in the real world, replete with all its biodiversity, ecosystem interactions and adaptation to the ambient chemistry. It allow researchers to exploit gradients in ocean chemistry that exist at sites such as hydrothermal or other CO_2 vent sites, across changes in pH among sites or with depth, or even between ocean basins (Barry et al., 2010).	(none)
Model output	Generation of data from models rather than actual field sampling or manipulation experiments.	(none)

ESSDD
8, 1–21, 2015
OA metadata template
L.-Q. Jiang et al.