Earth Syst. Sci. Data Discuss., 8, C26–C28, 2015
www.earth-syst-sci-data-discuss.net/8/C26/2015/

Open Access Earth System Science Data Discussions

Interactive comment on "A metadata template for ocean acidification data" *by* L.-Q. Jiang et al.

A. Velo (Referee)

avelo@iim.csic.es

Received and published: 5 March 2015

Carefully defining a template for metadata it's a very important task for the community, as metadata availability has a very high impact in understanding the datasets. The value of the information provided by a dataset can be penalized to useless if you do not have the appropriate metadata, and highly enhanced with a proper one. Thus, the work done here by the authors involves a great responsibility in future value of OA datasets, and a huge work on clearly understanding the needs. Reading the manuscript, it's clear the big amount of work done by authors for understanding the specific needs both in the oceanography field of OA data and at the current state of network services and technologies.

On the weakness of the manuscript, from my point of view, I could appreciate the following:

First. Template availability: You provide a DOI for template download, which is great, but the link points to a PDF portfolio. I don't know exactly if PDF ISO includes portfolios, but when trying to open it with typical browsers, it suggest installing Adobe software and doesn't let you open the template otherwise. I think that there are a number of de-facto standard archive formats that would be preferable, as a simple ZIP, or even an html page with the three files linked there. My advice here is to keep things simple where you can. PDF portfolio adds an unnecessary layer of complexity and does not offer anything additional.

Second. Track of versions: As authors indicate in pag. 3 line 26, it's critical for metadata distinguishing dissimilar data sets. Cruise datasets sometimes got their content updated due to corrections from PI, quality controls or other issues. This can be seen for many cruises in CCHDO for example. I think that metadata has to, at least, univocally indicate the exact dataset to which they belong. A track list of changes (similar to those short sentences of version control systems or changelogs) should also be very desirable, as it could help potential users and systems to univocally identify which data they are working with, and their tracked changes. There are a lot of ways to identify a dataset file, but one of the most extended (for files in general) is the md5sum of the file itself (supplied .csv from the website, for example), as users can easily redo it and check, and the generated hash is unique for a particular file. It would be very important to include this hash or something like this in the metadata to ensure the correspondence with a particular version of the dataset. Having the hash, the track list of changes could be something like a table or XML tree with rows that include a numerically increased version number, a date, perhaps an author ID, a short changelog comment, and the hash itself. Last, and perhaps out of scope, having a DOI for each version of dataset instead the one for all dataset versions should also be desirable, and so the template could accommodate it for the future.

Third. Dates and units Dates included in the templates for filling should warned to be in ISO 8601 format (YYYY-MM-DD or YYYY-MM-DDThh:mm:ss[Z]) to avoid most month-

ESSDD
8, C26–C28, 2015

day and even year mistakes. Units: It could be desirable for the template to include recommendations on units (or scale/temp for pH) according to best practice guides on the field. This can be out of scope, as the main goal of metadata is provide additional information about a particular dataset, but can help data submitters to standardize future data.

Fourth. Standardization – Interoperability I haven't seen any reference to SeaDataNet along the manuscript. SeadataNet is a "Pan-European infrastructure for ocean and marine data management", and provides metadata formats and vocabulary for different kinds of marine services. Their website has a section where they define some metadata formats (http://www.seadatanet.org/Standards-Software/Metadata-formats), providing the description of XML formats (ISO 19115 / 19139 compliant also) and their XSD schemas for validating/definition. CSR (Cruise summary report) and CDI (Common data index) could match with some sections of this OA metadata template. My question here is about interoperability from the viewpoint of data exchange along the international research community. I agree that having all data defined in XML ISO formats (and, of course, providing their XSD) greatly enhances this, as XSLT/XSL transformations can then be used to perform the conversion between formats (where possible), thus allowing the broad usage of data and its international availability. My doubt here is if authors have considered the existence of these SeaDataNet formats and vocabulary in the development of the OA template, as having planned this could ease these transformation and thus data distribution and discovery.

Interactive comment on Earth Syst. Sci. Data Discuss., 8, 1, 2015.