

Working Group Report for ASM 2012 **Ocean Acidification Data Management**

Organized by M. Gastil-Buhl, Lydia Kapsenberg*, Peter Edmunds, Bob Carpenter, and Carol Blanchette. (*presenter)

For abstract please see: <http://asm2012.lternet.edu/working-groups/ocean-acidification-oa-data-management>

At least 18 people attended our working group for 2 hours on Thursday afternoon, the last session of the week. Seven investigators, four information managers, six graduate students and an NSF representative participated. Five LTER sites were represented: SBC, MCR, VCR, GCE and LUQ.

Lydia Kapsenberg presented background material, the motivation and purpose of this working group, and then led us through four participatory exercises: two as a whole group and two in small groups. Our data will be archived within the LTER Network Data Catalog, not as a separate database. Agreement on terminology will ensure this important type of data is discoverable and compatible.

As a whole, we amended then unanimously approved the proposed definitions of data levels as follows, with the clipped and flagged data being the lowest level that should be posted to the public catalog while archiving lower levels:

- Level 0: raw .dat files
- Level 1: calibrated
- Level 2: clipped and flagged
- Level 3: aggregated, smoothed

Core Area

We discussed which of the five LTER Core Research Areas our data is best pigeon-holed but had to conclude that it does vary with datasets because ocean acidification (OA) is often measured in conjunction with biotic measures. Six investigators were able to point to example data already collected which fell into each of the five core areas, with the most falling into Movement of Inorganic Matter, and the least in Movement of Organic Matter. Seawater carbonate chemistry measurements belong in the Inorganic Matter area. If or when the network catalog offers a browse interface to datasets sorted by core area, there will be some OA datasets in each of the core areas. Each dataset must be categorized individually and a given dataset may pertain to more than one core area.

Search Keywords

A list of candidate search keywords from existing and anticipated datasets was used as a starting point for each sub-group to add terms, identify synonyms, narrower and broader terms. We asked 'what would your search for' and 'what words would you expect to find your own datasets'? These have been compiled and sorted. John Porter set up a clone of the LTER Controlled Vocabulary for Keywords for us to enter our new words and relationships and test their efficacy. Some of the proposed additions will be vetted by the Vocab Working Group but some must be staged for until after new data is submitted, since keywords are not allowed until at least one dataset contains them. We agreed that to ensure a dataset appears in a general query for this type of data, the term "ocean acidification" should be included as a keyword.

Measurements

It was unanimously agreed that we prefer a dictionary type lookup or reference rather than dictated standard names and constraints on types of measurements. The exact mechanism is yet to be determined, but the requirement is to be able to key each OA measurement to a set of predefined measurements within LTER and other research groups doing OA science such as European Project on Ocean Acidification (EPOCA) and the National Oceanic and Atmospheric Administration (NOAA).

Measurement characterization has long been discussed among LTER information managers but is still planned as a future effort. To introduce the concept, an example list of 63 measurements from the EPOCA data catalog were reviewed by four sub-groups for the purpose of seeing how a group outside of LTER that has been doing OA research longer has standardized their measurement names. We were asked to comment on whether the definitions were adequate, the measurement names recognizable, and whether LTER researchers would use the same units.

We agreed to clearly demark direct measurements from calculated values and in some cases to provide measurements in two forms, such as per mass for physical oceanographers and per volume for biologists. We agreed the calculations, including software and parameters, are important to include in the methods.

Conclusions

It was agreed we all want to avoid redundant data submissions. Not only will all OA data be submitted directly to the LTER network data catalog, not a separate LTER database, but also any BCO-DMO datasets will be submitted to just one of the catalogs and only cross-referenced in the other. [Mechanism TBD; possibly BCO-DMO may become a member node of DataONE.]

Future Work

It was suggested we might apply for a scoping workshop to bring together representatives from the now-concluded EPOCA, NOAA, and BCO-DMO so that LTER may harmonize our vocabulary, measurement and method descriptions with those groups.

Specific action items include:

1. new vocabulary keywords and relationships - propose to Vocab wg
2. compose an Ecological Metadata Language (EML) template for SeaFET data
3. list exemplar datasets (after revising them to meet newly agreed standards)