MBON Data Management Workshop, 2016 June 7-9
Marine Science Institute, UC Santa Barbara

Executive Summary

A 2.5-day workshop was held at UC Santa Barbara Marine Science Institute to begin coordinated data management activities across the three NOPP demonstration Marine Biodiversity Observation Networks (MBON) - Sanctuaries, Santa Barbara Channel and Arctic MBON - each in progress with independent funding. Each MBON plans to produce datasets for demonstration software (e.g., a map, web catalog or decision-support tool), and for cost-effectiveness, each proposed to leverage existing, familiar data management capabilities. However, for data products to be broadly reusable and software to be easily adapted, data management activities should be coordinated, which is being accomplished with supplemental funds to the three projects. The objectives for supplement Year 1 are to understand and document potential customer needs, local systems and capabilities, and culminates with this workshop. Preliminary work by individual MBONs included applying a common template to individual data management plans, and developing local inventories of anticipated data sets (not included here). The overall workshop goals were
> (a) To understand the needs of potential MBON data users, and
> (b) Initiate a coordinated approach for the three demonstration MBONs.
Some aspects of this approach and workshop outputs, e.g., documentation of customer needs or a processing roadmap have broader utility than US-MBON coordination; for example they are also relevant to a BON-in-a-box.

There were three major workshop objectives: First, to identify examples of the needs of potential users, and begin documentation of data delivery mechanics. By no means should these examples be considered the only uses of MBON data; rather they serve as a starting point. During the workshop, the usage scenarios examined were (a) National Marine Sanctuary Condition Reports (Monterey Bay, completed in 2015); (b) the Ocean Biogeographic Information System (OBIS), a web-based access point to information about organism distribution and abundance which will serve as a conduit for MBON data in GEOBON; (c) the IOOS data catalog which will harvest data descriptions from all MBON projects, and d) Axiom Data Sciences, contracted by IOOS to manage data and produce maps of spatial/temporal binned values. The data services considered were (again, not limited to) aggregation and updating of relevant datasets, delivery of data via web services, and production of data-based or derived products (e.g., diversity indices).

The second objective was to outline a common processing and data management roadmap for the three MBONs by comparing processing streams for several data types among the projects to identify major functional elements. Identifying common processing or workflow elements fosters a common terminology for describing the sequence of data processing steps and leads to a more standardized view for those creating end-user software (e.g., in Objective 1). Three MBONs plan to produce diversity indices and related data for a wide variety of organism groups, necessitating multiple data processing streams. Obviously - and at a minimum - research reproducibility requires

documentation of processes. But further, transferability (a general goal of MBONs) would require an understanding of each data processing stream, including sample analysis, 3rd party or untrusted data sources, obligate code/architecture, plus the necessary skill sets for completing each step. A major component of Objective 2 is a graphic depiction of the major steps in MBON processing sequences, with definitions of inputs, outputs, processes and interfaces. Elements we identify are conceptually analogous to the NASA data processing levels from raw (L0) to fully processed (L4).

The third objective was to begin plans for selection and use of controlled vocabularies to describe measurements used by the three projects uniformly. Considerable further planning, preparation and discussion will be needed. The group agreed that the basic approach for the MBON community should be to adopt one or more existing vocabularies as is possible (or required), but to augment those with missing terms or contribute definitions, using established pathways for those processes (e.g., with CFConventions.org or Marine Metadata Interoperability, MMI). However, it has not yet been confirmed that existing vocabularies meet *all* MBON needs either in structure or content. If they do not, the group will develop a plan for extension or adaptation that could possibly require additional funding.

Overall, to facilitate the reuse of their data, MBONs will need to adhere to regular practices. This workshop began to outline the following responsibilities of each MBON (see documents for details):
1. Define and document MBON processing and packaging formats
2. Provide adequate metadata with all data
3. Establish an unambiguous set of measurements for MBON population studies
4. Make datasets available via web services by contributing them to appropriate repositories

The workshop products are three documents (one for each Objective, above), accompanied by this Executive Summary, concluding with a short list of Next Steps, participants list and glossary of abbreviations. The participants suggest the following schedule for completion of workshop products:

1. Receive PI comments on three documents plus Executive Summary [October 1, 2016]
2. Receive DMAC committee comments [October 1, 2016]
3. Final version archived [October 15, 2016]
4. DMAC committee should meet in the Fall 2016 to plan the next its steps toward interoperable MBON data Next steps may require supplemental funding, e.g., to
   a. Produce data sets (metadata, data and code) that can be broadly shared, and appropriate for multiple uses;
   b. Refine community standards when current practices are not mature or widely used;
   c. Identify additional best practices and data protocols where needed.