

901512 An Integrated MBARI Time-Series Program (currently 901304)

Rating: Likely to be supported for advancing the data integration aspect of this program

MT contact: Scholin

Management Team Feedback:

Thank you for working towards an overall goal of achieving a system that makes it possible to integrate measurements throughout the water column. We realize that there is still much work to do, but applaud the effort thus far. Working on the data integration aspect of the project, and gaining a full understanding of the challenges that will be faced on that front, is a welcomed development.

Developing and operationalizing the in-water elements of this system is also an exciting proposition that speaks directly to the vision put forth in the Technology Roadmap. It is also a prospect that needs very careful analysis to ensure we understand the totality of its implications. For example, routine use of LRAUV and Wave Gliders would likely impact DMO, and perhaps require acquisition of additional platforms; what assets and support staff do we need to put in place to actually achieve the full system, and what are the projected costs of operating in that mode? Ramping up use of video and acoustic sensors brings many challenges with data storage, processing and search.

We know that the project leads understand these challenges. Our advice is thus to work towards a comprehensive understanding of the system requirements—from sensor/platform to data dissemination—and the operational realities that will bring, before charging ahead with individual proposals to implement various pieces of the system under a variety of project headings. Clearly some pilot scale experiments are needed and continued development efforts are warranted; the challenge is balancing a long-term plan that leads us to a sustainable system against the desire to move as quickly as possible towards this new mode of observation in different parts of the water column. This will take time—we will need to work together to bring this to fruition.

Because this project will be transitioning from a feasibility study to a full project, Mike Pinto asks that you use the new project number, 901512, for the Phase 1 proposal and budget entry.

As you prepare your proposal please consider the following:

- Provide an overall systems summary of the integrated time series effort so we can see how the various pieces fit together (hardware, software, data storage and query)we think this overview may be best captured in this Integrated Time-Series Program proposal where each subproject associated with the integrated time series effort is called out by its project number. Consider creating a table and/or graphic to capture the constellation of projects that when combined depict the overall integrated time series development effort.
- We note that many on-going and planned projects will generate an enormous amount of data. However, nowhere in MBARI's portfolio of projects—including

this one—do we see a concerted effort to scope the implications and needs for data storage, processing, and query, particularly with the intent that many systems will operate for extended periods of time (i.e., years). Many of the projects—alone and especially in combination—will overwhelm MBARI’s data storage/analysis cyberinfrastructure capabilities. In the past MBARI has largely dealt with this challenge by fragmenting data systems that serve individual research teams (“one project, one server”). This solution has inadvertently led to many problems with maintaining a host of individual systems at MBARI for different teams, and that approach is not optimal as we look to the future. For this reason we think it is in the collective interests of all to fully understand the data storage/processing/query implications, and the means by which we can mitigate potential impacts/costs that will put us on a sustainable data management path. To that end, the data Technical Advisory Group (data TAG, led by Kevin Gomes) has been asked to lead an effort to scope this problem and, working with the science/engineering teams, advance an objective analysis in 2015 about how we as an institution can accommodate the coming data deluge. Therefore, please include in your proposal an analysis of your projected data storage/processing/query needs so the data TAG can help us identify an appropriate, cyberinfrastructure strategy.

- As you develop the integrated time series concept, avoid a solution that leads to “rear-view mirror perspective” where the focus lies on recording changes in as many state variables as possible to enable retrospective analyses. This system should inherently include an effort to improve predictive capability and test hypotheses, too. With that in mind, is the integrated time series work really a series of targeted biological process experiments, or an attempt to gain comprehensive understanding of as much of the ecosystem as possible?
- What are the near-term science objectives with the integrated time series program, and how will that improve our ability to disseminate information and understanding to groups outside of MBARI that may use the knowledge for resource management and/or conservation purposes?
- As the data from the various programs merges and grows, we think it will be important to accommodate a machine analysis interface that can help point out complex relationships and highlight potential cause and effect interactions that would not necessarily be obvious to domain experts using the planned, manual data query interface. This is particularly the case for exploring genomic data and its relation to ecosystem structure and function; the addition of other data streams like video, acoustic and optical sensors will complicate this picture (no pun intended) further. How do we get ahead of this curve? One suggestion is to consider partnering with a signals discovery group such as that at the Pacific Northwest National Laboratory (see <http://signatures.pnnl.gov>). One of our newest Board members, Francine Berman, is also a resource to look to on this front.