Software and Data Systems Strategic Plan

2005.03.18

MBARI is well positioned to provide leadership to the oceanographic community in Data Systems technologies. As reflected in the MBARI strategic plan, our ability to collect data has outstripped our capability to understand it.

Focus on solutions for technology-limited marine-science research

In accordance with the institution's strategic plan, MBARI Engineering should focus on opportunities to advance technology-limited science, including both development of new technology and the integration and adaptation of externally developed technologies to the special needs of oceanography.

One group of data systems-related opportunities relates MBARI's longstanding efforts to develop long-term ocean observing systems. There is a fundamental need for standard interfaces to allow interoperability of instruments and data systems. MBARI can also make significant contributions by developing solutions to the logging, real-time distribution, archiving and dissemination of observatory data streams. There is a great need for understanding the appropriate time and space sampling scales for various oceanographic processes. In many cases, real-time decision making for observatories or field programs requires an improved integration of modeling with observations. Finally, effective automated quality control processes are required to deal with very large continuous data streams.

Several important opportunities exist for MBARI to contribute to how scientists analyze and interpret data. MBARI is unique in that we have more data flowing in per scientist than any other institution. This means we have a great opportunity to capitalize on the clear need for improved data mining and data query algorithms. Similarly, considerable improvements can be made in the techniques by which diverse data types are integrated into useful data products as well as integrating some new data sets, like acoustics, that do not have any good data management infrastructure. Also, data archives from multiple institutions can be integrated so that products derive from all of the archives without duplicating data between the archives.

The emergence of successful artificial intelligence technology opens opportunities to apply this to oceanographic problems. Adding an intelligent or adaptive capability can improve data quality control tools. Adaptive analysis of real-time observatory data relative to historical archives and data from other platforms may reveal previously unknown patterns and relationships. Event detection and response applications are required to meet the goals of the observatory programs. Automated interpretation of video data could enhance use of the data and provide for automated control of vehicles.

MBARI has been working on data problems and solutions in the oceanographic setting for a long time. The organization built many solutions for both observing systems and domain-specification applications, and has learned from this experience. Our work must take advantage of the skills and experience of the entire organization.

By focusing on the solutions to the most relevant problems, we can help avoid the threats of barriers to collaboration and inability to leverage external resources, as we will be addressing problems that interest the whole community. This will tend to foster willing collaborators who can bring resources to the collaboration.

If the correct and realistic focus for work is articulated, it will directly mitigate some of the weaknesses associated with the organizational dynamics and lack of focus that we have experienced in the past. A realistic assessment of what it will take to accomplish solutions is key, as well as a long term commitment to projects that span multiple years. The focus needs to be well defined so that both science and engineering are in agreement that the solutions built are valuable both internally and externally. If solutions address the concerns of the external community more than the internal MBARI community, there should be an explicit agreement that both science and engineering are aware of this fact and are still willing

to make the commitment. Within SDS (and the engineering group as a whole) there should be consistent processes and practices to enable effective execution of projects.

Investment in software and data systems can lead to decreased organizational costs through increased efficiency and more products (science) from the same observations (data) at small incremental costs. This is in distinction to developments of new platforms or new instruments, which often require additional costs to the organization once transferred to operational status.

Insist on teamwork.

The SDS can build teams internally with all kinds of experts (operators, developers, technicians, scientists, and outreach specialists) needed for creating and operating the systems, and using and evaluating the data produced by the systems. Leveraging this institutional advantage is critical for producing working, near-optimal comprehensive solutions that in fact solve most software and data system problems, for MBARI and the community. To achieve this goal, we must create processes which routinely enable these teams to collaborate effectively and reward those collaborations.

Ensure optimum scientific use of our assets and capabilities.

MBARI has one of the most comprehensive and diverse data sets for any region in the world (and access to even more data in collaborating institutions). We have more data per scientist than any other institution and our existing data sets include both unique data types, and most data types (via MOOS and MARS) that will be generated by large observing systems. We can and should be an exemplar of data integration for ocean observing systems, just using our own data internally. *Add something about external data/users.*

The oceanographic community is increasingly focused on observing systems such as those championed by the MBARI MOOS project. The NSF and other funding agencies recognize the importance of data systems problems to the community. MBARI operates (or will operate) and derives data from most of the classes of observing platforms used within oceanography (e.g. ROVs, AUVs, moorings, cabled observatories), and so efforts to work with MBARI's large, diverse datasets are directly relevant to the problems faced by the oceanographic community as a whole.

Complement, rather than duplicate.

By making sure SDS's solutions complement other work and efforts, we can help minimize the threat of competition and would help build our reputation as a contributor to the solution of difficult problems. In order to effectively do this, often the SDS group's contribution may be to integrate existing commercial and oceanographic solutions in unique ways. This requires that the group stay abreast of existing and starting developments in the commercial and oceanographic communities to make sure we are identifying the missing pieces of the solution correctly. During this work, we must avoid the threats of evolving standards and technology by knowing what those are early in the process and also, at times, help develop or steer those efforts.

Our solution focus should be on products that complement other developments, as this will make the most efficient use of our resources, which are limited. In addition, if we can stay involved in both commercial and external developments, we can more effectively collaborate on projects and get more accomplished by utilizing external resources to create solutions. Depending on the focus areas, we must bring in the required skill set to effectively provide solutions (some example areas are artificial intelligence, database and data management technology, and modeling).

Demonstrate the worth of our efforts.

In order to help improve our reputation, mitigate competition, and have an influence on evolving standards and technology, it is crucial that we demonstrate our solutions, their importance and their effectiveness. This will obviously help improve our reputation, but will also assist in the capturing of mindshare of other institutions and funding agencies. This will help remove the intentional and unintentional barriers of competition and will allow us to show MBARI's relevance by providing opportunities to show how these solutions fit into our overall roadmap. The key is to provide professional, consistent and unified demonstrations that build on previous efforts to show how we can effectively manage solution development by having structured internal processes and strong teamwork. This will also help us to have an effect on the evolution of standards and technology by giving us the opportunity to lead and participate in the teams steering those evolutions.

Promote effective communication, cooperation, and COLLABORATION.

MBARI's long-term and stable funding enables approaches and attitudes not available to other institutions. We can uniquely apply that funding by hiring highly qualified engineers, supporting relatively high risk or delayed payoff projects, and leading community-scale projects with relatively little direct benefit (monetary or staff size) to MBARI. By being open to community-based approaches (not limited by a not-invented-here attitude), SDS will extend MBARI's reputation and collaborative opportunities.

The ocean technology community benefits from and welcomes leadership on truly collaborative projects. Many technologists and developers at commercial and academic institutions are eager to partner on real, interesting problems. Through community collaborations, MBARI is capable of serving as an integration house for both point (specific) technologies and broad initiatives (e.g., observing systems, system of systems architectures, standards and protocols). MBARI is already involved with most or the entire major ocean observing initiatives.

Promoting effective communication, cooperation and COLLABORATION will be one of the most effective areas at alleviating all the threats the face the SDS group and the work we undertake. If we can effectively promote communication, cooperation, and collaboration, it will:

1. Simplify the management of competition
2. Improve our reputation
3. Help us procure and manage external resources
4. Allow us to participate, and stay abreast of, existing and new community standards and technology
5. Help avoid the unintentional misuse of our data and technology by helping the external community with how to effectively use and cite them.

In order to be effective in these areas it will take resources and require being pro-active. In a very similar manner to the demonstration of our efforts, our interactions in these areas must present MBARI, and our work, in a professional, consistent and uniform manner (our small size could be an advantage in this area).

Effective communication, cooperation and collaboration will help mitigate weakness in organizational dynamics and limited resources within the SDS group, engineering, MBARI and externally as well. Effective communication and cooperation will help keep us from losing focus, keep us from duplicating our efforts, and will promote teamwork. If teamwork and communication are working well, then with collaboration, we can mitigate limitations of resources by drawing help from other groups, other divisions, and from organizations external to MBARI.

Ensure optimum scientific use of our assets and capabilities.

Our current resources available in the areas of adjuncts, visiting staff, post-docs, research assistants, and interns, could be an effective outlet to help avoid threats and could touch on many, if not all, of the threats we face. Since these are, in some part, external resources, they can be effective in helping us stay abreast of competitive efforts, evolving standards and technology, and can also help improve our reputation by their interactions with their external contacts. Since they are also potential users of our data and technology, they can greatly help reduce the risk of their unintentional misuse because they will already be familiar with them.

By ensuring the optimum use of our assets, we will help mitigate the weakness in organizational dynamics and limited resources. If we can use our own assets to their fullest extent, it will help ensure that our focus is on the right problems and that the solutions we are working towards will have an impact on internal and external users. Also, if we make sure we are using our assets fully, that will help to ensure we have utilized all of our limited resource before we are declaring that we need more.

We will strive to work with developers, scientists and operators who propose to develop new data sources to include appropriate systems to manage the data to mitigating the problem of collecting more data than we can manage.

Seek external advice on the quality of our work.

Seeking external advice can also help avoid some of the threats we face in the areas of competition, reputation, and evolution of community standards and technology. Even though this is seeking external advice on our work, if we can provide a professional presentation of our solutions and how we developed them, it will give the external community an impression that MBARI is quite capable of solving difficult problems in a professional way. This will obviously help in our reputation, but can also capture community mindshare, which will give us a competitive advantage and keep us involved in evolving community standards and technology.

Seeking external advice will help mitigate a weakness we have in our engineering processes and execution of projects. In order to effectively use external advice, we will need appropriate processes and structure to communicate effectively to them. They may also have knowledge that will keep our focus and solution in check with any external efforts or existing solutions that we can use to keep us from wasting valuable resources. External advice can also give us a realistic check on how our products are working in the community instead of trying to guess how we “think” our products are working. [W]

Review roadmap with external experts?

Expand our relationships with regional institutions.

The SDS team is well coupled to the major existing observing efforts (MARS, NEPTUNE, ORION, IOOS, MOOS) and several campaigns such as AOSN, PLUS, ASAP .... We have established ties to the Monterey Bay oceanographic community, the wider oceanographic community, and the international metadata community, as well as to non-oceanographic data management and software efforts like Silicon Valley, NASA AMES and JPL, the astronomy community, Universities, Cencoos, HP Labs, and Intel Research. We can and should leverage these connections into collaborations, proposals, and major data integration efforts.

Large external developments exist in industry (XML, grid computing, Cyberinfrastructure, semantic frameworks) that can be exploited in the oceanographic setting. Several MBARI engineers have connections to people and activities in Silicon Valley.

Transfer our successes to the marine research community.

MBARI has an institutional commitment to disseminate our data and solutions. This can be used as the basis of a comprehensive data management approach. In addition to ready access to and publication of core stream data, MBARI can integrate even more regional data sets by embracing the enlightened self-interest of the science teams, who will certainly appreciate the value of their data and solutions being widely available in an integrated and professional way.

By efficiently and quickly transferring our successes to the research community we will avoid some of the threats of competition, reputation, and data and technology misuse. If the transfer is professional and smooth and the solutions solve real and focused problems, that bolsters our competitive advantage as well as improves our reputation. It can also reduce the risk of misuse if we can inform the research community on how to effectively use our data and solutions.

By transferring our successes to the marine research community we can help mitigate the weakness we have in engineering processes and with limited resources. In order to effectively transfer successes, we will need to have professional engineering processes in place that are up to par and equal with the community we are transferring the solutions to. Also, if our successes are used externally, it helps with limited resources in that they can help give feedback and provide further development on these products.

Communicate with the public.

MBARI can use its in-house and closely related education and outreach facilities to build academic, commercial, regulatory and personal collaborations, and to enhance our visibility and reputation. We must develop a focused plan, in collaboration with those programs, to maximize the benefit of our education and outreach resources.

Communicating effectively with the public can have an enormous impact on our reputation and can help avoid some of the threats we face in that area. If we can effectively use outlets like the Aquarium to implement and demonstrate interesting data and solutions to the public, that would directly reflect well on MBARI. Through communication with the public, we can, at some level, help mitigate our limited resources. If we are effectively communicating our vision and capabilities to the external community, we will grab more mindshare, which will better position us for external resources from funding agencies and programs.

Interactions with the Engineering Groups (Platforms, Instruments, Support)

Systems engineering issue? How do we interact? How do we integrate our activities?

CRITERIA FOR SUCCESSFUL IMPLEMENTATION:

As is stated in MBARI's strategic plan, finding criterion for successful implementation can be very difficult. Instead of proposing a type of metric to measure success at the end of each year, we propose the metric be a parallel effort that be kept in synch with our constantly reviewed and evolving roadmap for the SDS group. If we can track other efforts as well as our own and how that fits in to the various places on our roadmap, which should give us a good idea of how successful our work has been to date. We should see our solutions pop-up more and more on the roadmap if we are successful. This "results/status map" (think paved/unpaved roads) would be a place where we can track all significant efforts, publications, developments, etc. Also to ensure success, we feel we must avoid the "fire-and-forget" transfer of solutions and data to the community by staying in touch with those we have transferred our results to.

How do we measure that we have mitigated our weaknesses?

Metrics for successful.

Opportunities Text

Observatory Data Logging and Archiving

Solutions for the logging, real-time distribution, archiving, and dissemination of data from ocean observing systems, including long-term observatories, are an essential component of platform and observatory development efforts. MBARI has invested considerable effort into handling the data from cabled observatories, moorings, and other platforms, and is thus in a good position to complete and disseminate solutions to the community.

Integrating Modeling with Observations

Many oceanographic topics involve complex phenomena and systems that are being investigated using large and diverse datasets, and modeling is an essential component of the science. However, no satisfactory system exists for connecting real-time data to models, or vice-versa.

Integrating Diverse Data Types into Useful Data Products

MBARI produces and archives a diverse group of large datasets. In many cases, integrated views or visualizations derived from disparate data enable new insights, new correlations between the data and new science. In order to take full advantage of our data, MBARI needs to develop interfaces to the data that integrate all types (e.g. a common portal for all datasets using consistent interfaces for each) and provide access to useful data products.

These data interfaces would have to touch on not only observatory type data (data streams), but also discretely sampled data like frame grabs, cruise samples, laboratory samples and sample analysis results. The discretely sampled data also brings about another opportunity of sample curation. In order to make effective queries and use of the discretely sampled items, a curation system could be built to ensure the longevity of the samples, much like a museum. Often times sampled data is lost when a scientist is no longer associated with them and a great historical and reusable data set is lost.

Data Mining and Query Algorithms

MBARI, and the rest of the oceanographic community hold vast and diverse stores of data but use crude methods to find and extract information relevant to any particular topic. MBARI has an infrastructure (in SSDS) that can be used as a basis for developing more advanced, useful, and easy to use systems. MBARI can also take advantage of the considerable progress in this topic that has been made external to oceanography.

Integrating Data Archives from Multiple Institutions

MBARI has considerable data complementary to other institutions' holdings and of interest to the oceanographic community. However, MBARI, like most other institutions, operates its data archives with little linkage to external archives. The increasing Internet bandwidth available now makes it possible to integrate data archives from multiple institutions so that data products derive from all of the sources without a need for duplicating data between the archives. Developing these sorts systems will require standard system interface protocols and application of semantic computational tools now starting to mature (e.g. Marine Metadata Interoperability effort). Some of the needed infrastructure software has been developed externally, but only limited useful implementations exist. There are also areas of data collection that do not have good management solutions, like acoustics, that we could integrate as well. MBARI also has some reference software (VARS, SSDS, possibly SamplesDB) that can take advantage of existing or proposed interface protocols.

Integrating data archives throughout MBARI and making it widely available.
Leverage internal data management talent?

Developing Standard Interfaces for Interoperability

MBARI has invested considerable effort into developing standard interfaces to allow interoperability of instruments and data systems (ISI and SIAM and SSDS). Interoperability is necessary in order to facilitate multi-institution use of high cost platforms and observatories.

Developing Quality Control

Effective quality control processes are necessary for handling vast data streams deriving from long-term observatories. MBARI needs to develop capabilities for integrated, automate real-time quality control checks on data streams. Tools for manual quality control are also required, including improved interfaces for identifying and flagging errors for future users.

e.g. rock chemistry quality (Lamont petdb), seamount database

Understanding Optimum Sampling Scales in Time and Space

MBARI holds datasets from many platforms that sample processes over a variety of time and space scales. These datasets are appropriate for developing optimal strategies for using our limited platform resources.

Automated Data Interpretation

Advances in Artificial Intelligence technology external to oceanography provide an opportunity for MBARI to advance automated data interpretation techniques. Use of vision technology for automated interpretation of video data could enhance use of the data and provide for automated control of vehicles (ROV and AUV). There is also interest in real-time analysis of observatory data relative to historical archives and data from other platforms. The community's long term goals for observatories include oceanographic event detection and response applications. Finally, there is a need for adaptive quality control software.

An emerging deep ocean technology is ocean observatories, where instruments of all kinds, including video cameras, are deployed for long periods on the ocean bottom or in the water column, enabling fine scale temporal sampling. The instruments are powered from shore or water surface sources, with data streams to shore over network cables or surface-deployed wireless communications links. Systems such as the proposed plate scale NEPTUNE observatory seek to deploy multiple video cameras to watch science experiments, and continuously observe biological activity and changes in species kinds and abundance. However, for these cameras to be practical and cost effective scientific tools, automation will need to be developed to flag and classify events in the video streams, as aids to science users, and technology for event response (such as zooming a paired close-up camera onto a potential event of interest) will need to be developed and deployed. “Research on pelagic community structure represents one of the largest instrumentation and intellectual challenges for an RCO. This type of research will require instruments that can be used to identify the organisms that are present on a variety of scales and some elements of their dynamics (growth, physiology, ecology).” [1]

Virtual Observatories (“Point-and-click Oceanography”)

Echoing the astronomical community’s efforts towards Virtual Observatories (“Point and Click Astronomy”), advances in Cyberinfrastructure along with Ocean Observatories enables an oceanographic data avalanche. Breakthroughs in cabled and moored observatory platforms, sensors, and computer technology allow ocean scientists to produce terabytes of data and images. These datasets will cover the ocean from different perspectives – overhead satellites, surface observations, coastal arrays, and deep sea installations, all arriving to shore in near real time. In a few years it will be easier to "dial-up" a part of the ocean than wait many months to access a UNOLS ship. “Point-and-click oceanography” could greatly

[1] http://www.geoprose.com/cabled_wksp/pdfs/reconn_rpt.3.2.04.pdf

enhance the scientific output from the expensive platforms and instruments by bringing more PI eyes onto the same data sets. With the advent of inexpensive storage technologies and the availability of high-speed networks, the concept of multi-terabyte on-line databases interoperating seamlessly is no longer outlandish. More and more catalogs will be interlinked, query engines will become more and more sophisticated, and the research results from on-line data will be just as rich as that from "real" ocean observations. These technological developments will fundamentally change the way oceanography is done. These changes will have dramatic effects on the sociology of oceanography itself. *(Shamelessly adapted from the NSF National virtual Observatory project).*