Science Panel Comments and initial Q&A
WHOI and NCEAS Data Management Concept Papers
July 11, 2011 Draft

Overview

The science panel has received the concept papers from NCEAS and WHOI. Both groups are willing to collaborate with Axiom and perhaps even each other; they are both highly experienced and all indications are they are capable of producing a relevant and usable database.

Both groups are very well-regarded by all panel members and there is confidence in their abilities and products overall. We are pleased they are interested in the Council's programs, past and future data.

Below are individual science panel comments on the concept papers, followed by an initial Q&A.

Review of Axiom's Work

Axiom has submitted their late work to the Council. NCEAS staff, as well as an independent IT contractor, have recently reviewed Axiom's past and proposed work and they uniformly came back with a fairly positive review, though they note the proposed work is not implemented, and implementation is a considerable step.

Individual Science comments are also cautiously positive about Axiom completing their work and receiving solid reviews on their product but are also supportive of Axiom's local involvement, for example:

" I have concern about [not having a collaboration with Axiom] because I think that collaborating with Molly and AOOS is important in providing a long-term home (and an appropriate one) for the data. I also was impressed by the testimonials included in the revised Herring and Long-term Monitoring proposals from the bench scientists (at least the leaders) in support of the AK team."

1. One Panel member, who is familiar with Axiom's recent work, notes the following:

The pros' of Axiom include 1) local knowledge - currently, Axiom is working with AOOS and the PWS Science Center. So, they are working with folks that may have similar needs and uses for the data. 2) If the historical information is going to be included, the lead at Axiom has a history with the old data and would have a jump on how it was formatted, stored etc. 3) The interface they use is Google Earth and they've recently upgraded some software that makes their system more easily transferrable among other companies. 4) Their data presentation allows for manipulation and actual use of the data; it's not just spreadsheets of numbers. 5) They seem to be relatively good at bidding their projects, and don't come in over budget at the end.

Many of the 'cons' associated with them are related to growing pains of a new company, Axiom went from 2 - 8 employees relatively quickly, so sometimes they bite off more than they can chew. And, given the time frame required, deadlines can be missed. Service folks have learned to manage this aspect by requiring shorter, interim steps during the development of a project. But as they grow, if the company management doesn't keep up with their rate of getting big projects, then time management may continue to be an issue.

Prior Experience of NCEAS and WHOI

1. The WHOI group has broad experience and has basically done the same thing at least twice before. There is no doubt that they will successfully pull the majority of the data together for posterity in an organized and consistent way that experts will be able to extract what they need from.

2. WHOI has experience managing big, oceanographic programs. If they were to be directed to work without Axiom, unsure of how much the lack of local knowledge/contacts would affect work.

3. The Woods Hole Group has true credibility and experience with biophysical coupled systems, which is what climate change driving of biological systems calls for.

Risk vs. Reward: The Potential for Synthesis

1. NCEAS is proposing to reorganize the data with an ecological metadata approach that will make it easier for many people to use the data in many different ways, which will greatly increase its value. The WHOI approach is well tested and essentially risk free. The NCEAS approach is riskier, but has the potential to generate a lot more benefits. A lot of risks would be taken with the synthesis part of the program, but most of them paid off with impressive deliverables in other projects. It is not easy to get scientists to step outside their training and work together, but it is not impossible either.

2. NCEAS would provide the "added value" of some structure and focus on the "what does it all mean ecologically?" which will provide the means for adaptive management of the monitoring program and the resources themselves in the future. Otherwise, you run a risk of saving money and ending up with well-managed data collection for the sake of long-term datasets.

3. If all of this can be provided by either NCEAS or WHOI, then my preference from a perspective of broader access to the data and its potential interpretation is to go for NCEAS proposal with the higher risk and possibility of higher reward as described by [another panel member].

4. My take on the two groups is that both are well-respected, established entities that could handle the data management task for the projects. To me, it seemed clear that BCO-DMO was quite concerned if they went with Option 2 ($1,437,632) in that this funding level could be a potential problem down the road. As such, my read was that from their perspective, Option 1 ($2,060,110) was the one to go with, or the one they felt more certain of. In that sense (unless I am misreading), BCO-DMO is not necessarily a bargain.

   I am certainly not an expert in different cyberinfrastructure systems, but it is my understanding (from colleagues who use it) that the KNB repository network that NCEAS utilizes links a large number of entities and data sets and it is considered essentially an “open source” system. The advantage of this is open access (low or no cost) and therefore interest by scientists outside of the immediate EVOSTC umbrella who want to utilize the data should be high (of course at the appropriate time when the data should have permitted access). While in the short term (<10 years) it is probably the Alaska scientists and environmental managers that need the most access to the data (including historical data sets), the legacy of EVOSTC science programs will ultimately be based on the use of the data at national/international levels, where peer reviewed publications emerge highlighting the long term data in PWS and the GOA, and ultimately establishing the program and the science funded from the EVOSTC as a classic model for oil spill damage assessment and restoration.

It is my view that this is where BCO-DMO and NCEAS differ significantly. While BCO-DMO will work with scientists collecting data and assist with analyses, the NCEAS proposal to have academic working groups take on analyses **and** synthesis of the data will result in high impact scientific publications at a national/international level (they have a strong track record here). Therefore the bang for the buck can be quite high. This is a very unique aspect of NCEAS and is something that does not exist with other data management groups, to my knowledge. As we discussed at our last meeting, postdoctoral fellows are a "cheap" and effective means (they have the motivation) for publishing synthetic papers that EVOSTC scientists are all part of but may not have the time to synthesize across disciplines and projects.

NCEAS is the "risky" choice I suppose, although they have an excellent track record so I am not sure how much of a risk exists. It is worth having a heart to heart with Director Ed McCauley regarding NCEAS' future. While NSF funding is shifting away from them, they are funded by UCSB and have numerous other non-NSF funding. Frankly, all of these data management entities depend on funding from other projects that can expand or shrink (as many are now), and in a sense, are all on "soft" money.

5. I truly do agree with [another panel member's comment] that an open source is the ethical way to go with the AK data sets after a period of exclusive access by the APIs and students has expired, so that aspect of the NCEAS proposal has appeal.

Implementation: Consultation, Oversight and Management. *ED comment: The contract with either entity would be with the TC and would include oversight of Axiom, if applicable.*

1. In order not to lose even more money on overly optimistic estimates of time and money needed to deliver data management products, set up a series of well thought out and specific milestones for the funded entity. Hold back of payments should be considered as a project management option.

2. Which of these two entities is most likely to enable an eventual broad-scale synthesis of the Gulf of Alaska? In other words, how well will each of the competing products mesh with other efforts in the Gulf of Alaska and Bering Sea; suggest consulting with NOAAA, NPRB, UAF, etc.

3. Whoever is awarded this contract should, at a minimum, provide review and oversight to inform the Trustees and their staff whether or not data management goals and objectives for each project are being met on the agreed upon timeline. This could include a review of all historical EVOS datasets (and any other the science panel chooses to specify) for a "second opinion" as to which of these have sufficient metadata or other value to archive and be made available since Axiom has already proposed to review historical datasets.

Implementation phase: Integrating future and past datasets *ED comment: Data management would begin with the PI meeting this fall to examine future datasets so they are accommodated appropriately. Past datasets would be examined by a working group, including managing agencies, to determine suitability for inclusion before investing in integration.*

1. Data collected from routine surveys, collected consistently in time and space, are usually far more valuable than from short-term, unique studies or experimental work, whether in the lab or the field. To be valuable

the data must be recorded in a form where it can be accessible – and also there must be a narrative that explains the nuances of the data, including its weaknesses. Both competing proposals appear to provide that but the WHOI may have a better grip on this aspect.

2. In general, most of the herring projects probably will be relatively short-term. The main value will be in the report(s) produced by the PI's with very limited value of the data for long-term monitoring. Clearly the data should be recorded and made accessible, but it may be a challenge to organize some types of herring data in a format compatible with the electronic forms suggested by NCEAS. This comment, however, does not necessarily apply to the other collection of projects (benthic monitoring, oceanographic surveys, etc.).

3. Even if one or both proposals were fully implemented there still are issues that warrant attention. One is the status of data from the past 20 years of EVOSTC-funded work. Will these data ever be made accessible – in a format compatible to the data archives that are now under consideration? The other issue concerns related data from other agencies, especially ADFG, and perhaps even NOAA. Some essential data are not available, except perhaps by special request. As an example, the raw herring data (lengths, weights, ages) collected over time in PWS are not available – except in summary form. Many researchers do not want data summaries; they need the real stuff.

Initial Q&A from Panel:

What sort of investment in hardware are you proposing for Axiom or your group for the long-term work with/storage for EVOSTC?

From WHOI: WHOI is unclear as to what Axiom's role has been to date and what may have been accomplished thus far with respect to the EVOSTC data. We have not received any documentation regarding Axiom, and we are therefore unable to guess at how we might interact with them now or in the future. In general, we work well with others, identifying cost-effective, efficient collaborations and opportunities for complementary work whenever possible. We are certainly open to working with Axiom, but we know nothing about the organization at this point.

The budget we put together includes a dedicated server and data storage for the EVOSTC data in the first five years. If EVOSTC wanted the data transferred (for whatever reason), we would likely transfer the entire server with all required software (all public domain, no licensed products), to the recipient as identified by the Council, or back to the Council for distribution. As mentioned previously, the data would already have been contributed to NODC for long term archive. I anticipate that we'll do that routinely as we do for other data managed by BCO-DMO.

From NCEAS:

We are currently placing an approximately $250K order for data archival hardware, as one of the original nodes on the NSF DataNet DataONE project (http://dataone.org). This would be relevant to our interactions with EVOSTC, relative to data archiving solutions. The NCEAS concept paper is not intended as a "full" proposal specifying hardware and backup, etc. Rather, it was viewed as a critical and highly beneficial complement to what we imagined Axiom or WHOI might be providing in terms of more baseline cyberinfrastructure and informatics solutions.

With regard to staff, NCEAS staff has been working on on some separate initiatives relating to environmental investigations of the BP Oil Spill in the Gulf of Mexico.

Financial Stability

From NCEAS:

The NCEAS organization has the full support of the UCSB Vice Chancellor for Research, and the NCEAS' Ecoinformatics Group is an established research unit on the UCSB campus.

From WHOI: WHOI is a private, non-profit corporation, not a public entity. In March 2011, Standard & Poor's declared their outlook to be 'stable', with an AA rating based on:

•role as the world's largest private oceanographic research and higher education institution with a mix of federal grants from the National Science Foundation (NSF), Office of Naval Research (ONR), and the National Oceanic and Atmospheric Administration (NOAA);

•Balanced operations on an unaudited full accrual basis for fiscal year-end Dec. 31, 2010, demonstrated by a $2.2 million surplus and an unrestricted net asset increase of $3.9 million;

•balance sheet based on unaudited fiscal 2010 results, with cash and investments of $361 million equal to 180% of adjusted operating results and 584% of outstanding debt;

• Lack of current debt plans, existing fixed-rate debt structure with essentially flat debt service and low maximum annual debt service (MADS) of 2.3 % of unaudited fiscal 2010 adjusted operating expenses;

• Recent fundraising success with the completion of a $200 million capital campaign in 2008;

• Growth in the size of sponsored research programs to $176 million or 37% since fiscal 2007.

If the entity were to be compromised in the future, what would their suggestion be for the EVOSTC data?

6. Panel Comment: The WHOI proposal involves some major investment in essential hardware, which the NCEAS one does not. All the EVOS data is not going to get stored forever on four laptops. Perhaps NCEAS already has the hardware available from other sources, but if not someone needs to look closely at their long term plans.

From WHOI: While there are very few groups who offer management the way we do, there are several options for long term archive of data. As mentioned in the documentation, all data managed by BCO-DMO are archived at NODC for long term preservation. NODC is considered a 100 year archive and is compliant with the ISO OAIS model for information archive. It is true that if our infrastructure, WHOI and the other BCO-DMO funded projects, disappeared, we could not do the management of the EVOSTC data for the budget we have supplied. We are taking advantage of a lot of synergy, being located at WHOI and there is an economy of scale when managing data from many projects. The data are fully backed up, and could be relocated to another server if WHOI were to be compromised with respect to cyber-infrastructure. A hurricane or fire is the most likely scenario, and the data are protected should either or both occur.