SeaWiFS Data System Critical Design Review

Mark R. Abbott and Robert Evans

16 April 1992

We have several concerns that were not adequately addressed at the SeaWiFS Data System CDR. The concerns revolve around the areas of overall system coordination, understanding of the requirements, and system implementation.

System Coordination

1) The overall data system is dispersed over four distinct, independent groups: ground system/mission operations, calibration/validation, data processing, and data distribution (which is the responsibility of the Version 0 GSFC DAAC). This has resulted in a disjointed system where the responsibilities and interfaces are only vague. In many cases, there were apparent conflicts as to who was responsible for which functions. This was most obvious in the SeaWiFS/ DAAC interface. Even within the SeaWiFS project, there was a lack of coordination in terms of the GSFC HRPT data; it appeared that the cal/val group was responsible for data processing although there was no apparent mechanism to ensure that the cal/val processing was compatible with the primary SeaWiFS processing system.

We recommend that the project immediately implement a program to ensure compatibility between the various components so that they appear as a unified, seamless system.

2) Much of the data flow between components is handled as file transfers (such as navigation). While such an approach may work, it is likely that incompatible file formats, timing issues, etc. may appear in the total system. A fragmented approach is more prone to such breakdowns simply through misunderstandings and lack of communication. There appears to be little accountability **between** the components.

3) The components appear to operate autonomously, even at the budget level. It appears that each component has its own budget and is free to spend resources as it sees fit. For example, the mission

operations component has spent an inordinate amount of time and money on navigation which is a well-understood problem. It is **not** a challenging technical problem. Such an expenditure may be well within the mission ops budget, but other issues, such as HRPT level 2/3 processing support, may be ignored for lack of resources within the cal/val group (which apparently has responsibility for this activity). If in fact the cal/val group does the processing for HRPT/LAC data to be used for the cal/val process, it is not apparent that the programs and procedures used will be identical to those used by the GAC data processing group. As such, there is a distinct possibility that the application of the cal/val results might not be consistently applied in the two processing groups. Without close, structured management at the top, these disparities may result in a seriously unbalanced system. Similarly, a particular issue may be ignored because each component feels its the responsibility of another component. For example, the Japanese OCTS mission represents a unique opportunity for intercalibration of multiple ocean color sensors. Yet the OCTS mission is ground station-limited, a problem that has not been communicated to the appropriate people. Also, no one has examined the possibilities of joint data processing or data sharing because this is viewed **entirely** as a cal/val problem.

4) The MODIS project is responsible for funding much of the SeaWiFS algorithm development work. The MODIS project is also confronting many of the same algorithm integration and data processing issues as SeaWiFS. There is clearly an opportunity here for cross-fertilization of ideas and experiences between the two projects. With the replacement of MODIS-T by a follow-on SeaWiFS mission (EOS-COLOR), the need to coordinate SeaWiFS and MODIS activities has taken on more importance. However, the SeaWiFS project has no specific plan to interact with MODIS.

Understanding of Requirements

1) One of the key lessons learned during the CZCS era was the need to provide access to the level 2/3 processing software. The present project plan is simply to provide documentation. This is totally inadequate. The project must provide documented, tested, up to data code to users and HRPT station operators. On the other hand, this does not mean that the project must provide support. Rather, the hardware configuration and the appropriate source code must be available to any interested person.

2) Along with the need for access to software, another critical lesson learned from CZCS was the need to develop portability, resilient software. The project must not become **beholden** to a particular vendor. This will not only ensure that the project retains flexibility to take advantage of new advances in technology, but also that the user community is not forced into a particular configuration. Along with developing portable code, the project must also develop a mechanism to **test** the software in other environments. Portability is not ensured simply by writing software to particular "standards." A notable exception to portability is the reliance on the Silicon Graphics GL graphics library for image display. This is **not** a standard. It should be recognized that the SeaWiFS processing for the GAC data involves two major components, level conversion processing and the overall process/data management. The first component must be portable across a representative selection of vendor equipment. This portion of the overall processing system is sufficient to permit the HRPT stations to produce products consistent with those produced by the project. The second component is not required by the HRPT stations and can be tailored to project requirements.

3) The project has not developed a baseline set of requirements to describe how the user community will interact with the total system. There were numerous examples of this deficiency. There are no end-to-end data delivery requirements. That is, how long will it take for a user to receive a final, "polished" version of the data based on the "refined" ancillary data fields? The ability to process and deliver 'refined' products critically depends on availability of validated ancillary data fields. The project is dependent on the interagency data MOU's functioning in an efficient manner. Headquarters and project management should carefully review the existing mechanisms to insure timely availability of these fields. The project seems to view its only responsibility for data delivery ends with delivery to the DAAC. How long will it take for HRPT station operators to receive orbital elements and ancillary data fields as well as software updates?

There are only vague requirements concerning HRPT stations. The present baseline only includes HRPT stations as data collection and distribution sites, not processing sites. However, the user community expects to receive such level 2/3 data. The presentation on the second day helped alleviate our concerns, but there is too much left "unsaid." For example, the encryption problem was not adequately addressed, and it was not clear who was responsible: OSC or GSFC. The HRPT issue needs a separate and thorough review. In particular, this review is needed in a time frame that provides an adequate time period for the HRPT station operators to acquire funding and purchase to necessary equipment (e.g. the equipment necessary to support the decryption and data storage equipment required by the embargo period).

There are no models for how users will interact with the system. Nothing has been written down. Users are counting on SeaWiFS yet the data ordering/delivery system has not been described. Both the project and the DAAC seemed to assume that the other group would provide the data searching/ordering capabilities; at least the resolution of this issue was not clear from the information presented during the CDR. More attention has been paid to data processing but not in how users will search, order, and receive data. The present system appears to be a copy of the CZCS system, which will not be adequate. No thought has been given to automated orders; the only model is a person interacting with the system. The project views this as a DAAC responsibility, but the overall goals of the system should be to support science, not deliver data to the DAAC. The present user model is woefully inadequate.

Algorithm and data submittal by users is another user interaction that has not been considered. Given the EOS data policy, it appears that such interactions will be required, yet the project is not prepared to handle them.

Lastly, there has no thought given to user metrics to monitor system performance and user satisfaction. Such a plan should include both automated and manual (questionnaires) monitoring systems. The system is designed as a one-way, data delivery system.

System Implementation

1) The present method for data delivery from the frame formatter to the data processing system relies on IEEE cards. Why not take advantage of existing hardware and software that handles guaranteed data delivery, such as TCP/IP and Ethernet? The IEEE approach is outdated and unreliable. Protocols are much more effective. The use of the IEEE bus electrically ties the 'frame formatter' machine to the data processing machines and subjects them to whatever electrical transients (external power spikes or

lightning) that potentially would be coupled to the formatter by the external communications circuit.

2) There is absolutely no need for VGX graphics capabilities. This will allow the project to do 3-D rendering, and other graphics functions that are not needed. No graphics are needed in the data processing systems. Resources would be better used to increase the number of data display sites, not provide redundant graphics capability as the limiting step is the number of people who can view and quality-control the data. As graphics are needed, the capability should be provided by low cost/high capability graphics stations. Given the number of files that require review, more stations are likely to be needed than the single display provided by the VGX tied to the main cal/val processing machine. A reason that this approach was rejected concerned the need for additional software licenses which made this approach costly. Functions such as database access could be provided by the main server while the graphics station need only use the remote display capabilities provided by the vendor.

3) The present Silicon Graphics systems should not be upgraded to 8 processors; Digital Review has shown that doubling the number of processors from 4 to 8 increases the performance by about 15-20%. A four processor machine delivers throughput equivalent to ~3.5 processors while an 8 processor machine delivers the equivalent of ~5.5 processors.

4) An upgrade to multiple R4000 machines will be not be painless; Silicon Graphics has announced that this upgrade will require a complete change to the back plane, and the present binary compatibility in the SGI line will cease.

5) How will data be filled in if the frame formatter goes out of synch?

6) There is a single point of failure between the control processors (VAX's) and the data processors (SGI) at the router. The project must provide an Ethernet connection between the two systems so that there is a redundant link between the Ethernet (VAX) system and the FDDI (SGI) system.

7) There is no plan to prototype or test and evaluate new hardware or the various builds. Such a plan must be developed.

8) The plan presented for hardware redundancy included use of multiple processors. However, the only provision for sharing of disk or other peripheral resources required physically moving cables between machines. This is a high risk approach with opportunity for catastrophic affects (physical or electrical) to the equipment. Several vendors offer equipment that will permit the SCSI devices to be directly connected to multiple computers thus requiring only a software reconfiguration and not a physical rearrangement of the equipment.

Summary

Given our concerns, we recommend that the project undertake another CDR in the near future to address these issues.