REVIEW COMMITTEE SUMMARY[1] REPORT on:

Data Management and Archiving System (DMAS)

Preliminary Design Review (PDR)

for

Management − NEPTUNE Canada and VENUS

Submitted by Ian Barrodale (Review Committee Chair)

August 11, 2006

Identification and brief biographies of the seven Review Committee members:

Richard Baldwin is a geoscientist at the Pacific Geoscience Centre and has managed the western computer operations of the Canadian National Seismic Network since its inception.

Ian Barrodale is President of Barrodale Computing Services Ltd. (BCS); in 2004 BCS delivered a 235-page report to VENUS and NEPTUNE Management on DMAS needs and requirements. He was the Founding Chair of the Computer Science Department at the University of Victoria.

Bob Keeley is in the Integrated Science Data Management Directorate, Department of Fisheries and Oceans. He is the chair of the joint IOC/WMO JCOMM Data Management Program Area and a member of the steering teams of the U.S. DMAC and European Union SeaDataNet projects.

Paul Ross is currently managing a development and integration team delivering solutions for a large BC Health Authority. For the past year he has been managing an enterprise integration project based on a Service Oriented and Event Driven Architectural approach.

Jörg Sander is an Assistant Professor of Computing Science at the University of Alberta. His current research interests include spatial and spatio-temporal databases, as well as knowledge discovery in databases, especially clustering and data mining in spatial and biological data sets.

Randall Sobie is a Principal Investigator for the Research Computing Facility at the University of Victoria. He is currently investigating how Canadian researchers will use the Grid to store and analyze the vast data sets from the current and future experiments.

Bill St. Arnaud is Senior Director Advanced Networks for CANARIE Inc., Canada's Advanced Internet Development Organization (www.canarie.ca).

---

[1] This report summarizes an earlier Review Committee report delivered to Management on July 4, 2006.

Executive Summary

This Summary Report is based on information gained from documentation supplied to the Review Committee, formal presentations made by members of the DMAS Team at an all-day Preliminary Design Review (PDR) on June 23, 2006, responses by the DMAS Team at the PDR to questions and observations from the Committee, subsequent oral and email submissions, and subsequent email interactions between the Committee members.

At the concluding closed session on June 23, the Review Committee first recorded its joint recommendations and observations, and then assigned relevant topics to each Review Committee member in order to provide additional guidance to the DMAS Team, and to address issues raised both in the supplied PDR document and during the meeting/presentations. A Review Committee report (dated June 30, 2006) was presented to Management on July 4 containing recommendations and observations, partial answers to questions that were raised in the supplied PDR documentation, a brief summary, and an appendix containing detailed supporting material from Committee members, as well as a submission from the VENUS Science Director. The purpose of the present report is to summarize the findings in the June 30 report.

The Review Committee made ten joint recommendations and observations during the closed session at the PDR (listed on page 3 of the June 30 report to Management), and so they are reproduced in the next section of this report. In addition, eighteen other recommendations and observations gleaned from the Appendix of the June 30 report are included in the next section. As is appropriate at a preliminary design stage, some of this content is supportive of work accomplished to date and some is somewhat critical. Clearly, most of the issues raised here with notes of concern must be dealt with before the Critical Design Review takes place towards the end of 2006.

Recommendations and Observations of the Review Committee

1) It is essential that a DMAS Steering Committee exists, composed of end users and technical representatives, and that it reports directly to NEPTUNE Management. See also 20) below.

2) Use cases must be documented and translated into technical requirements, such as network bandwidth, system availability, replication/sharing of database, etc. See also 17) below.

3) Human resources for quality assurance (QA) of software should increase from one to two or three, and appropriate effort must be devoted to developing relevant test cases and benchmarks. See also 18) below.

4) The Committee noted that there was very little discussion on the relative merits of the current choice of file server plus database as compared to a database alone. The Committee also noted an absence of justifications for choices made to date. Such justifications must be provided in the Critical Design Review. See also 11), 18) and 27) below.

5) There needs to be some clarification on source code versioning versus database versioning. The metadata also need to be extended.

6) The Committee generally approved of the web services approach for scalability, handling metadata, and controlling instrumentation. See also 13) and 18) below.

7) Because of the use of some leading edge technology and the relative lack of experience of current staff, the Committee supports setting aside a portion of funding for external consulting help.

8) The PDR documentation provided did not address in any way direct real-time feeds such as seismic feeds (via the bus etc.). See also 19) and 24) below.

9) The Committee is aware that this is a Preliminary Design Review, encourages the DMAS Team to focus on objectives to be met over the next six months, and notes that the Critical Design Review occurring in December 2006 will be a very important milestone prior to deployment in 2007.

10) The Committee urges the DMAS Team and NEPTUNE Management to address the question "whenever NEPTUNE goes live, will the *first version* of the DMAS meet the *actual requirements*?" Too many software projects over-promise and under-deliver, so take due notice of what users want to do, but be careful what you promise the DMAS will do.

11) The Committee urges the NEPTUNE DMAS Team to thoroughly explore available database product options, so that it can satisfactorily justify and defend its ultimate choice in the months ahead. Careful benchmarking, using relevant data and tasks, is absolutely essential in this regard.

12) The resource management issues are not yet well developed. In particular, the statement was made that high priority change requests could result in lower priority data acquisition to be turned off. It was implied that this might be handled in an automatic way. The rules for altering data acquisition schedules must be very clear to a data collector before they "sign on" or there will be unhappy users.

13) The DMAS Team reported that there has not yet been any benchmarking of how well a web services approach will be able to perform based on volume of the retrieval. The Committee believes that such benchmarking is important to inform users of what the performance characteristics will be of Service-Oriented Architecture (SOA)-based data retrieval.

14) The whole area of the data storage was not dealt with. This was noted in the document as out of scope of this review. The volumes of data that will be delivered by sensors and the success of how they are handled will be how the DMAS is measured. This must be covered in the Critical Design Review.

15) The acceptance of CANARIE CIIP funds required an SOA. The Enterprise Service Bus implementation of this architecture is, as stated in the presentation, relatively new technology. The Committee is concerned that no discussion of alternatives was forthcoming and finds it difficult to judge whether or not this is a reasonable choice. This is a critical element of the design and it *must* work.

16) The notion of a central power management facility needs to be extended to all resources that are shared. In particular, bandwidth needs to be managed so that no experiment can impinge on any other without the requisite authorization.

17) The Committee discussed the need to make sure that functional requirements and use cases rather than design documents form the basis of test plans and test cases. This is especially necessary for successful end-to-end testing to ensure that the project delivers value to the user community.

18) There are some concerns with an SOA design and specifically web services being prone to performance issues. QA needs to play an active role in testing and identifying potential bottlenecks before these are promoted to production environments. Within all components of the system special attention needs to be given to performance of complex data and streams as they move through the system. This is particularly true of web services and potentially the database and data storage system. It is recommended that performance targets be put in place and benchmark testing be performed on production hardware and software as

soon as it is feasible so that alternate approaches may be implemented if necessary.

19) The choice of an SOA for the communication with devices and data sources is perceived as a forward looking, appropriate design decision – assuming it can also meet possible real-time and high-throughput requirements (and assuming its usage can be successfully communicated to the science users). SOA presents a potentially scalable, flexible and robust technology that allows users to conveniently define complex and easily extendable operations on (and reactions to) the data streams produced by the instruments.

20) The Steering Committee needs to provide specific targets to the DMAS Team for the fall of 2007 (when Neptune becomes operational). It is likely that only a subset of the system is required in 2007; for example, the control of instruments could be restricted to expert users. The Steering Committee should identify the required functionality for 2007. Further, the data rates need to be specified so that the proper hardware can be put into place at both the shore station and the wet plant.

21) The Committee was concerned that the time between the Critical Design Review and the deployment of an operational system is approximately 9-12 months. This is an extremely short time for developing, prototyping and testing any solution.

22) In view of the emergence of various buoy and cabled ocean observatories around the world, if any prior related work does exist it would be useful to get their input on the software architectures employed.

23) There are several companies that make virtual firewalls, and the Committee recommends consideration of the use of virtual firewalls in front of every major process rather than using traditional domain or protocol-based firewalls.

24) There was no mention of any real time mirroring of data - so that external users can get a copy of the real time data feed, in addition to what is being sent to the DMAS. How will real time mirrored data be represented - who will control the flow, etc?

25) The Preliminary Design, as presented in the PDR document, had insufficient metrics or alternative designs to prove that the selected design as presented would meet/exceed the performance requirements of either VENUS or NEPTUNE.

26) In the opinion of the VENUS Science Director, it was not obvious that the summaries and recommendations of the 2004 BCS report had been followed up on, adopted, or refuted. He stated, for the record, that VENUS felt the BCS report/work was valuable, and that VENUS would have liked its recommendations to have been either more influential or at least considered in the DMAS design.

27) The VENUS Science Director also conveyed his concern to the Review Committee that the present Relational Database Management System (RDBMS), which is Sybase, is clearly insufficient, as demonstrated during early data flow. It could not handle the "ingestion" of the rather simple and "small" (10kB/minute) data stream from a single ADCP, and so this rather fundamental oceanographic data type (currents) was then labelled "complex" and relegated to flat file status, and as such would not be directly "searchable". The Committee noted that these two points are of crucial concern, and they must be resolved (if necessary through involvement of external consultants) before the Critical Design Review.

28) Finally, a Management issue brought (in confidence) to the attention of the Review Committee Chair suggests that prompt attention should be paid to some current DMAS human resource concerns. Successfully achieving DMAS project timelines will require a cohesive team effort.

Postscript

The members of the Review Committee appreciated the opportunity to provide technical and other advice to the DMAS Team and Management of VENUS and NEPTUNE Canada. These are very exciting and challenging seabed observatory projects, and the Review Committee's aim was to provide objective input that will encourage development of the best possible Data Management and Archiving System. We wish you well in your efforts to achieve your goals.