DMAS Action Items in Response to the Recommendations of the PDR Review Panel

The following document is a response to the various recommendations expressed in the Preliminary Design Review Panel report as received on August 11, 2006. The report followed the review itself, which took place on June 23, 2006. In addition to a narrative addressing each recommendation, the table below provides scheduled dates for various actions that are associated with the recommendations of the Panel.

The DMAS team would also like to express its gratitude toward the work performed by the review panel members. The input we have received is extremely valuable in that it is an indication of where we are perceived to perform adequately, where progress must be made to refine our plan, where care should be exercised and where the communication aspects of what we are doing should be emphasized.

Action items.

Action	***Date***
Provide a clarification note on database and code versioning.	September 10, 2006
Recruitment of an extra QA person	September 30, 2006
Produce a charge and mandate for a Steering Committee.	September 30, 2006
Identify Steering Committee Membership	September 30, 2006
Produce an RFP for the database/datawarehouse system	October 6, 2006
Organize a first Steering Committee meeting	November 10, 2006
Organize a science workshop to obtain high-level DMAS requirements	November 10, 2006
Organize a Critical Design Review	December 20, 2006

Recommendation 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.*

The DMAS team agrees with this recommendation. We will work with the NEPTUNE Canada co-chief scientists, the VENUS project scientist and the Project Directors to prepare terms of reference/mandate by end-September 2006 and to assemble a steering committee a month later. The Committee composition will balance science user representatives and technical experts.

Recommendation 2.

> *Use Cases must be documented and translated into technical requirements, such as network bandwidth, system availability, replication/sharing of database, etc.*

An effort will be made to assemble use cases that will translate into quantitative and qualitative requirements for the system we are building. A science workshop involving the current stakeholders and getting inspiration from the best visionaries in this area is scheduled for November 2006. The work will be completed by early November and will be presented at the December critical design review.

Recommendation 3.

> *Human resources for QA of software should increase from one to two or three, and appropriate effort must be devoted to developing relevant test cases and benchmarks.*

A new job description has been drafted and recruitment is scheduled for September 2006.

Recommendation 4.

> *The Panel 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 Panel also noted an absence of justifications for choices made to date. Such justifications must be provided in the Critical Design Review.*

The DMAS team acknowledges the request for justifications as far as the technological choices. Those will be explicitly detailed at the critical design review.

Recommendation 5.

> *There needs to be some clarification on versioning, source code versioning versus database versioning. The metadata also need to be extended.*

By versioning, we mean the data description structures that will enable the storage of *historical information* about the infrastructure (instruments, locations) as well as what happened to them (failures, repairs, etc.). Clearly we not only need the supporting structure, but also the software code that will be history aware.
Source code versioning is clearly a different topic. We obviously have source code versioning, supported as we are by tools such as "subversion".
The DMAS team is not clear as to what extension of the metadata the review team is alluding to.

Recommendation 6.

> *The Panel generally approved of the Web Services approach for scalability, handling metadata, and controlling instrumentation.*

The DMAS team acknowledges this comment.

Recommendation 7.

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

The current DMAS staff has an accumulated experience amounting to about a hundred years of experience in all matters related to databases, archives, user interfaces, sensor interfacing etc. Nevertheless, several $100K are already put aside for consulting help in the area of enterprise service bus, thanks to the CANARIE grant. Moreover, some money for database consultancy is also available and can be supplemented if need be. As noted in the PDR document, there is already a contractual arrangement in place with HIA/NRC for external advice on selected issues.

Recommendation 8.

> *The PDR documentation provided did not address in any way direct real-time feeds such as seismic feeds (via the bus etc.).*

The issue of handling of real time feeds is still at the definition stage. We are currently contemplating a scheme that will make use of the publisher-subscriber mechanisms that we are putting in place. A full presentation of the design and support for direct real-time feeds will take

place at the CDR and will in particular address the seismic data transport and interfacing issues.

Recommendation 9.

> *The Panel 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.*

The DMAS Team is taking good note of that comment and is fully aware of the challenge that this represents.

Recommendation 10.

> *The Panel 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.*

The DMAS team members are acutely aware of this danger and will strive to produce an infrastructure that will deliver core functionality and satisfy key requirements (such as data acquisition and data delivery). With a solid core infrastructure, we will be in a position to add features for the future stages of developments and turn new requirements into simple development projects that can cheaply be plugged into the system. However expectations have been raised so high by the visionary scientists leading this project that it will be difficult not to disappoint some.

Recommendation 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.*

The DMAS plan of work has a task and a milestone scheduled for the Fall of 2006, which addresses this exact point. An RFP will come out in September, which will invite database vendors to propose their solution to address our data storage needs. It is our hope that the outcome of the study will be available by the CDR.

Recommendation 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.*

Due to the capabilities of the infrastructure, it is not anticipated that resource management will be a priority at start-up. However the system needs to support the future seamless implementation of resource management. Resource management decisions to allow or deny access to say, power resources at a particular location and time will be automatic and based on resource availability and experiment priority. The NEPTUNE Canada Science Advisory Committee (SAC) will be the body recommending the priority of individual experiments in need of significant amount of resources. Clearly, users will be informed about the status of their instrument and the associated risk.

Recommendation 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.

Benchmarking will take place as soon as we have the first representative sets of web services in place, scheduled for the end of this year. Performance and data access expediency are key drivers when delivering a service. We should be in a position to report on this topic by the CDR.

Recommendation 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.

There are two aspects to the data storage in our design. The first deals with the data from scalar sensors, which we intend to leave in a fast data warehouse system for rapid access and easy cross-correlation. The second is the complex data types produced by multi-dimensional sensors and which do not have requirements related to content search.

The first issue will be dealt with through our RFP to database vendors which will take place before the CDR. (See recommendation 11). As to the complex file storage, the VENUS project received as in-kind donation from the NRC a complete software package to manage large amounts of files of arbitrary sizes. This system is in use at NRC's Herzberg Institute for Astrophysics and is successfully handling close to a 100 TB of scientific data. More information on it will be available at the time of the CDR.

Recommendation 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.

The review committee is right to be concerned by this issue. However, as explained in our description of the risks and their possible mitigation, we have mitigated the risk by working with IBM software and their relevant experts. While this approach does not eliminate the risk, we believe that partnering with an industry leader reduces it to a reasonable level. Moreover, we should not forget that this approach is the one favoured for the implementation of the US ORION Oceans Observatory Initiative, a ~$30M US cyber-infrastructure project.

Recommendation 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.

The DMAS team is fully aware of this fact. Clearly bandwidth may be an issue in the future and it will need to be dealt with as well. Management of bandwidth will, at lower levels, be handled by the network managers supplied with the transmission equipment and switches.

Recommendation 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.

The DMAS team is aware that testing should be made against the requirements and this is indeed what will be implemented.

Recommendation 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.*

Performance (and general behaviour) of the key elements of our design is fundamental. In that respect and for the message queue system, our expected efficiency numbers are being provided to IBM so as to scale the system appropriately. Web services will be closely monitored as to their performance and in case of trouble with some key ones, a different approach could be chosen. For the database, our current model is being assessed and as mentioned before, will be the topic of an RFP coming up soon. Finally, the DMAS QA has been tasked with the performance assessment of the critical software elements that we are deploying.

Recommendation 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.*

DMAS appreciates the vote of confidence expressed with this comment and, as already mentioned in response to points 13 and 18 above, will strive to gauge the scalability of the approach.

Recommendation 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.*

This recommendation will be in the charter of the Steering Committee, as defined in recommendation 1.

Recommendation 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.*

The DMAS Team is very conscious of that fact. but the team is making the most of the opportunities presented by being able to test many of the concepts it plans to suggest to use for NEPTUNE Canada on the VENUS observatory, thereby learning a lot and figuring out solutions that will work and identifying those that will fail.

Recommendation 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.*

The DMAS Team has a representative on the ORION Cyber Infrastructure (CI) committee. This committee is composed of representatives from most of the significant IT players in the area of Ocean science. The solutions that we are proposing (e.g., the use of an SOA) are also those that are considered for many of the current on-going initiatives, most notably the upcoming NSF Oceans Observatory Initiatives' CI implementation.

Recommendation 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.*

The virtual firewall concept, one where a firewall "engine" runs on the same platform as the one the user intends to protect, is an interesting proposition. It is likely easily implemented in a web services environment. We will investigate that option.

Recommendation 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?*

If mirroring here is meant to represent access to real-time data by external users, we would like to argue that this will be provided by the "publish and subscribe" architecture that we are proposing to use. Users in critical need of real-time data will be able to subscribe to them as close as possible to the instrument whereas those with less critical needs can subscribe to the stream from the data centre, through the web services that we will make available.

Recommendation 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.*

The proposed architecture is indeed very new and can bring some uncertainty with it. However, it will be tested on the VENUS observatory first, as part of the mandate we received from CANARIE and we will use that experiment to extrapolate up to NEPTUNE. Also, as the critical elements of the system rest on technology provided by a strong commercial partner which has ensured us of the feasibility of the approach, we are confident that it will satisfy the needs of the observatory.

Recommendation 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.*

The Barrodale report was useful in that it contained a good summary of the types of instruments to be expected on the Observatories. It also did a good job in advising us about the current data standards in geo-sciences in general. It reviewed in a useful manner the approach followed by other science data centres. It also went on suggesting implementation possibilities and data storage technologies available today. However, the report did not propose recommendations per se. It provided (to quote the document itself): "[...] an extensive compendium of information, background material, experience, and resources on which to base the development of a DMAS for VENUS and NEPTUNE".

Recommendation 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.

The DMAS team believes there is some misunderstanding here.

The initial concern about storing the ADCP data in Sybase was with the rapid growth of a data set that would never be used in the anticipated way: they were no immediate user requirements to query live ADCP data and needing instant access to any single value in the complex matrices returned by ADCPs and ZAPs. Therefore, to limit the risks and the overheads at this stage, it was suggested that the data set be stored on a less demanding resource.

Recommendation 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.

This is a serious issue, which the NEPTUNE management is fully aware of and is taking measures to remedy. These measures include a re-organization and change in the responsibilities within the team, and may include team-building initiatives, if necessary with third party assistance. This reflects the growing pains of a young organization.