NEPTUNE DMAS

Revolutionizing ocean science

NEPTUNE Canada
University of Victoria
PO Box 1700 STN CSC
Victoria BC V8W 2Y2
Canada

NEPTUNE Data Management and Archiving System (DMAS)

What is DMAS?

The Data Management and Archiving System for the NEPTUNE Canada project will assume the following roles:

- Collect, index and archive all data produced by the various instruments and transmitted back to shore by the network infrastructure, or "wet plant".
- Provide all users (scientists, educators, students and the public at large) with ways and means to search and access data from the instruments, be it newly acquired ("real time") measurements or older observations.
- Allow privileged users (scientists, instrument providers) to control the instruments in the water and provide ways and means to change data acquisition parameters.
- Ensure that neither commands sent to instruments nor autonomous instrument reactions will endanger the health and safety of all or part of the infrastructure by trying to acquire unreasonable amounts of resources (power and network bandwidth).

These roles represent the current top-level definition of the functions to be performed by the DMAS. They are quite varied, but all depend on a complete, up-to-date, permanent information repository about the entire network.

Assistant Director, Information Technology
Benoît Pirenne

January 2005

UVic

Voice: 1 250 472 5353
Fax: 1 250 472 5370
www.neptunecanada.ca
E-mail: bpirenne@uvic.ca

NEPTUNE DMAS

Basic, simplified key DMAS requirements

Relevant to the roles described above, the following requirements are key to the definition of the system, but are provided here as examples only and do not represent a complete set. The purpose of this list is to convey the challenges that the DMAS will have to address.

- DMAS will have a major role in implementing the data access policy of the observatory.
- DMAS shall, as much as possible, provide users with data in physical units. The system shall therefore offer on-line data calibration mechanisms.
- DMAS shall, at a minimum, be inter-operable with other Ocean Observatory systems and initiatives. DMAS shall therefore offer access interfaces resting on widely accepted standards.
- The DMAS shall buffer data at the Shore Station and/or rely on the instruments to do so to minimise the extent of data loss in the event of loss of communications between the DMAS and the Shore Station
- A provision to reduce the data volume at the Shore Station shall be implemented since the communication capacity between the shore station and the data archive and distribution centre is likely to have a bandwidth smaller than the maximum throughput of the "wet plant" network.
- The DMAS shall support many types of instruments, including "smart" instruments capable of considerable autonomous activity, and other very simplistic instruments that may not be "Observatory Aware".

DMAS Schedule

In order to provide a working system in time for the installation of the network infrastructure in 2007 and most of the scientific instruments in 2008, the following schedule must be respected:

- *User Requirements*: the first half of 2005 will be devoted to finalizing the User Requirements of DMAS which will rest on what we will learn from the following four elements:
  - o The development of a prototype system, which will implement some of the basic functional requirements,
  - o The adaptation of the prototype to the VENUS project needs,

Assistant Director, Information Technology
Benoît Pirenne

January 2005

NEPTUNE DMAS

- o The network infrastructure final design, as provided by the industry proposals in January 2005,
- o The initial science requirements in terms of instrument types and numbers, data volumes and real-time event detection requirements as expressed in the science proposals received in January 2005.

- *Functional Requirements*: By the end of Q2, 2005, based on the User Requirements identified using the elements above, we will be in a position to finalize Functional Requirements that define the System from an engineering and design standpoint.
- *Initial System Design*: An initial system design will be completed in early Q3, 2005, that complies with the Functional Requirements and the project goals.
- *Acquisition Plan*: On the basis of the Initial System Design, an Acquisition Plan will be prepared and implemented by the end of Q3, 2005, that details how implementation will be handled, which partners (industrial and academic) will be involved, how partners will be selected etc.
- *System Development*: between Q3 2005 and 2008, we expect the system to be undergoing development, to provide a number of intermediate deliveries that will be offered to VENUS, as ways to continuously improve on the interim system operated by this project.
- *Infrastructure control*: In 2007, in parallel with the deployment of the infrastructure, the interface between DMAS and the network's resource management should be put in place.
- *Initial Deployment*: We expect the first science instruments to be installed in 2007, with the majority of the instruments being deployed in 2008. With their arrival, data collection, interaction with the instruments shall commence together data delivery to users. This will signal the start of the operational phase of NEPTUNE.
- *Operational Phase*: DMAS will be in maintenance mode, following the evolution of the instruments, the computer technology and the infrastructure.

The DMAS Prototype

Motivation

As a first step towards better understanding the technical requirements of a DMAS applied to Ocean Observatory instruments, a small-scale prototype was developed based on

Assistant Director, Information Technology
Benoît Pirenne

January 2005

NEPTUNE DMAS

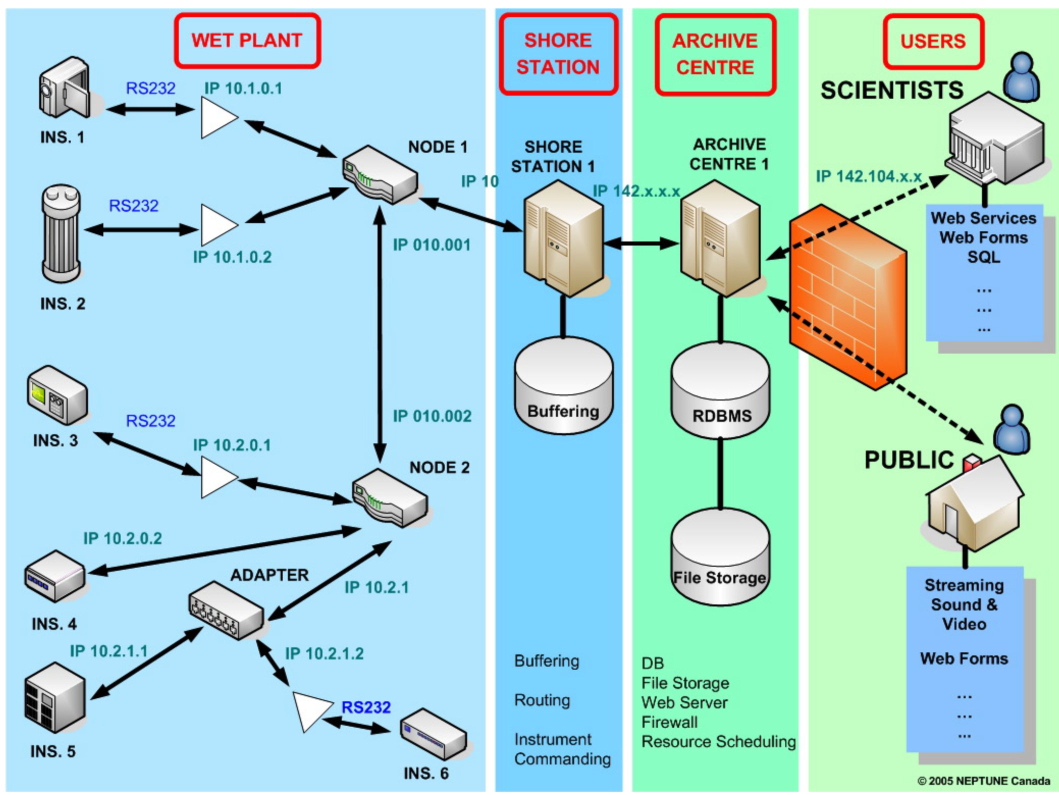
Assistant Director, Information Technology
Benoît Pirenne

January 2005

commercial off-the-shelf components. Much past experience in this area as well as competitive offers from commercial companies helped significantly.

The prototype intends to emulate the entire network with its key functionalities on a very small scale. The structure is represented in Figure 1 below and shows the three major elements of the system: the *Network*, the *Shore Station* and the *Data Centre*.

FIGURE 1: Conceptual overview of a simple ocean observatory from a DMAS perspective: it is composed of a "wet plant" with instruments and supporting nodes, a Shore Station with cable landing and a data centre, providing services directly to scientists and the public at large.



Another reason for having a DMAS prototype ready early, is its potential use for VENUS. As a matter of fact, VENUS is in a situation whereby its assets are ready to go in the water, but the corresponding data acquisition system is, for various reasons, essentially inexistent. This new deadline leaves no time to build a system using a formal, traditional approach of going through user and functional requirement, design and implementation phases. Moreover, the potential to learn from VENUS was very motivating

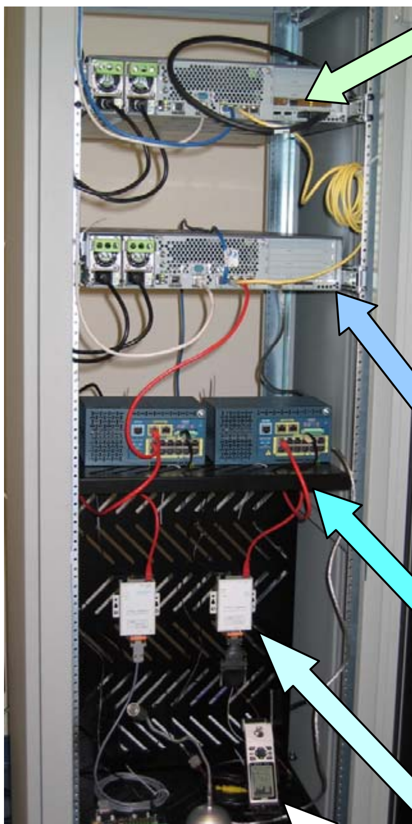
NEPTUNE DMAS

and therefore, a mutually beneficial situation was waiting to be exploited and responded to.

Prototype DMAS hardware

The prototype was implemented as "NEPTUNE in a rack" as showed in the table below. A description of the hardware and software used is provided.

Please note that all the commercial products mentioned here are provided for information only and do not represent commitments on the part of NEPTUNE Canada towards any of these vendors for the future, fully capable system.

Concept DMAS	Hardware	Software
	• *Data centre*: Sun V240, 2x1.25GHz, 8GB RAM, 2x73GB, 4x1000baseT	• *Data centre*: Sybase Adaptive Server Enterprise **RDBMS**, ESO/HIA's starcat: an old **database query and plotting tool**, ESO's WDB: **web interface** allowing navigation, query and retrieval of database entries, files retrieval; HIA's **file archiving and handling system**.
	• *Shore Station*: Sun V240, 2x1.25GHz, 4GB RAM, 2x73GB, 4x1000baseT	• *Shore Station*: one custom-developed program per instrument type, interfacing the instrument to the archive; file archiving client
	• *Nodes*: Two Cisco 2955 switches, connected to Shore Station in a loop	
	• *Serial to IP converters*: Moxa DE-311, TCP server configuration	
View of the various components of the hardware and software in the prototype DMAS system.	• *Test instruments*: Davis Vantage2 weather station, Garmin 45 GPS receiver, Ontrak ADR2205 relay board, microphone, PTC-201 steerable internet camera	

Assistant Director, Information Technology
Benoît Pirenne

January 2005

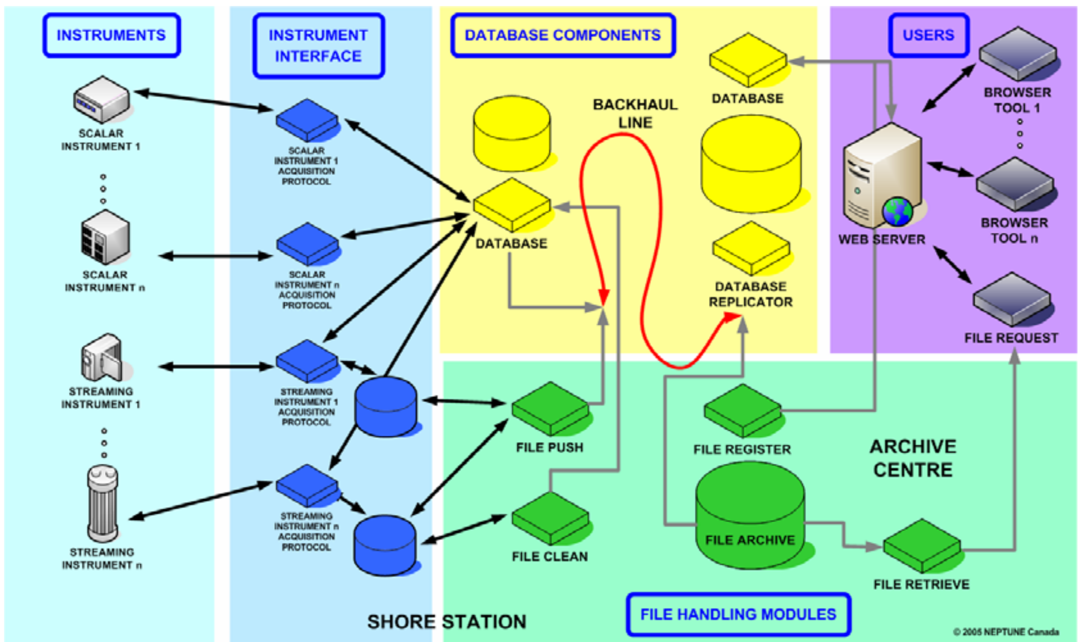
NEPTUNE DMAS

Prototype DMAS software

Structure

The overall system architecture, taking into account the requirements mentioned above calls for a number of components, as depicted in Figure 2 below. This structure allows for a potentially unreliable and low-capacity back-haul line, allows for event detection algorithms to be located at either the Shore Station or the data centre and assumes most users will get data from the data center. Not depicted here is the parallel channel for remote control of specific instruments in need of real-time piloting.

FIGURE 2: Functional view of the prototype system. See text above for details.



As mentioned before, a lot of the software components of the prototype are either well-known commercial off-the-shelf products, contributions from partner organization or from the public domain. The only software development that had to take place was -and presumably will be for the future system- in the area of interface between the system and the various instruments. As those instruments come from different vendors, they tend to have different characteristics, hardware interfaces and protocols for interacting with them.

An attempt at simplifying this interface can be seen in a NEPTUNE instrument requirement whereby all instruments should possess an Ethernet data communication interface.

Assistant Director, Information Technology
Benoît Pirenne

January 2005

NEPTUNE DMAS

Even if this is achieved, the protocol issue will not be solved and code will still be required to translate the "language" of each instrument type into a DMAS lingua franca. This particular problem can be divided into two parts: the first one deals with the basic communications language of the instrument (its command language) whereas the second one has to do with the interpretation of the instrument data.

The first problem has been elegantly solved by at least one commercial product, but the second one has no solutions yet known to us. The reason for that solution vacuum has to do with the nature of the instruments: they are mostly developed by commercial companies who consider their data format proprietary and tend to force their customers to use their own software to do the data analysis and visualization. If VENUS and NEPTUNE are to be observatories providing users with powerful cross-correlation capabilities, it will require a paradigm shift to storing data in a common form.

Data Acquisition

The principles under which this prototype operates for instrument reporting scalar values (physical measurements) are the following:

- Commands are sent to instruments to read their data on a regular basis (e.g., every 2 second)
- The output from an instrument is separated into its individual data items (e.g., temperature, pressure, salinity, etc.)
- If any of the values have changed since the last measurement, they will be tagged with a timestamp, a data type, and an instrument identification and will be recorded in the database

For streaming instruments (hydrophones, video cameras) or instruments producing large volumes of data with a complex structure (still cameras, multi-dimensional sounders, etc.), the principles are the following:

- Wait for user commands (including operation mode, duration of data collection, positioning, number of snapshots, etc.)
- Send command to position device and record data for specified duration
- Possibly process the file produced (e.g. compress)
- Record file metadata (start/stop dates, observing conditions)
- Make file available for storage

Assistant Director, Information Technology
Benoît Pirenne

January 2005

NEPTUNE DMAS

Data Retrieval

Data retrieval can be done thanks to special or general-purpose query tools that understand the SQL language and can interface to the database management systems. At present, a web-based query tool, using European Southern Observatory's public domain WDB web to database interface, is implemented and allows dynamic navigation through the database structure.

Data can also be loaded directly into popular tools such as MS-Excel through the use and proper configuration of Microsoft Query.

Data retrieval efficiency is extremely good with the system at hand and the "data warehouse" concept now implemented in this prototype. This prototype will address VENUS' data storage and retrieval requirements during VENUS' initial operation. However, the prototype's medium term scalability where random access to tens of billions of individual measurements will have to be offered, must be carefully studied with database technology vendors prior application to other projects such as NEPTUNE.

Event Detection and Reaction

Thanks to the simple database design, it is easy to write scripts that implement event detection algorithms. On a regular, unattended basis, the scripts query the database for the latest data of a certain type, run averages, find extremes, correlate values from different instruments, etc. Simple scripts have been implemented in the prototype to run queries against the database and find moments at which weather parameters reach a pre-defined threshold. When the conditions are met, any action such as sending email, turning on a relay, etc., can be triggered.

Thanks to the high performance of our database management system, response from multiple event detection code running simultaneously has been excellent, even in the presence of large amounts of data in the database, providing confidence in its ability to sustain the needs of VENUS.

Other event detection algorithms can also be conceived where the scientist will define intelligent software agents that will not query the database for data but rather plow through an incoming video or hydrophonic stream, run pattern recognition algorithms and react when particular things are "seen" or "heard" in the real-time feed. Reaction to such events could also include action on other instruments.

Assistant Director, Information Technology
Benoît Pirenne

January 2005

NEPTUNE DMAS

More generally, similar programs could be implemented whereby the software constantly monitors a video feed and produces say, fish counts per unit of time and records this new information in the database for long-term (seasonal, decadal) statistical analysis.

Supporting VENUS

The future of the prototype is intimately linked to VENUS: it will morph into the "Interim DMAS": The test instruments such as weather stations will be replaced with CTDs (conductivity, temperature, and depth water monitoring systems), chemical analyzers, microphones with hydrophones, and web cams with their underwater equivalents. This transition will take place during the spring of 2005, with gradual delivery of functionality in parallel with the installation of the VENUS infrastructure. More details will be provided as progress is made.

Final Thoughts

This initial prototype has been constructed as the first design phase in order to better comprehend the requirements and challenges of the future enhanced DMAS system; simultaneously meeting the immediate functional requirements of the VENUS project. These deliverables for VENUS have had to be accomplished within a three month timeframe, with the realization that this design will be modified to incorporate the evolving needs of both NEPTUNE and VENUS.

Consequently, the design choices selected for this phase of the prototype do not necessarily reflect the final choices of either project; but represent a basic, minimalist system that will serve the early needs of the VENUS project.

Benoît Pirenne

version 1.0

Last modified: Thu Feb 3 08:18:15 PST 2005

Re-Posted: Thursday, February 3, 2005

Assistant Director, Information Technology
Benoît Pirenne

January 2005