

CANARIE Intelligent Infrastructure Program



Statement of Work

**Toward a Service Oriented Architecture and Workflow
Management for VENUS and NEPTUNE**

**v. 1.4
April 12, 2006**

Cover Sheet

Project Number: CIIP - 19

Name of Project: Toward a Service Oriented Architecture and Workflow Management for VENUS and NEPTUNE

Project Goals:

The goals of this project are to provide the VENUS and NEPTUNE Canada Cabled Ocean Observatories with an integrated scientific instruments management, the capability to deliver event information to users, as well as integrated access to distributed compute and data resources through the use of innovative technologies. The wealth of new and old data will be easily exploitable through the use of workflow orchestration tools, existing grid processing infrastructures (including the high speed networks interconnecting them) and the underlying technologies related to web services. The expected results will include a rapid enrichment of the data archive, turning raw data into directly exploitable information much faster and in a more elegant way. The project is innovative in that it will provide an integrated approach in dealing with many different data sources (sensors and databases) while at the same time defining an approach to accelerate the extraction of knowledge from the data collected. Canada will rapidly benefit from the new architecture and management technologies that will be developed given the aggressive schedule of the project. Moreover, the country's profile abroad will be raised significantly as other organizations world-wide are contemplating the construction of similar ocean observatories and are watching very closely our progress.

Project Costs:

Direct Labour:	418,495.97	Contribution requested from
Overhead:	272,022.39	CANARIE: 1,099,433.83
Benefits	79,216.44	
Direct Materials:	20,000.00	
Special Purpose Equipment:	160,000.00	
Subcontractors/Consultants:	457,400.00	
Travel:	51,000	
Other:	16,000 (audit costs)	
Total Costs:	1,465,911.78	
Total Eligible Costs (all in-kind removed):	1,318,911.78	

Sources of Project Funding:

Working Capital:	219,477.94	Stacking (%): 75
Borrowed Capital:	0	
Other:	147,000.00	
*SR & ED Tax Credits:	0	
*Other Government Sources:	0	
*CANARIE:	1,099,433.83	

Total Funds: 1,465,911.78

Stacking (%) = the sum of * Categories expressed as a % of Total Eligible Costs

Contact Information:

Lead Contractor: University of Victoria
Address: 2300 McKenzie Ave. Victoria, BC, V8P 5C2

Type of organization: University

Contact person: Benoît Pirene
Phone: (250) 472 5353
Fax: (250) 472 5370
E-mail: bpirene@uvic.ca

List participants: IBM Canada
Address: 2d Flr - 1803 Douglas Street, Victoria, BC, V8T 5C3

Type of organization: For Profit

Contact person: Rob Heuchert
Phone: (250) 881 8671
E-mail rheuch@ca.ibm.com

Project Overview

An integral part of the VENUS and NEPTUNE Projects¹ along with instrumentation and the “wet plant”, the data management and archiving system (DMAS) will be in charge of the 24/7 data acquisition of a whole array of instruments and sensors, of their long term data storage and retrieval and of the network resources management.

The varied nature and the large number of fixed and mobile sensors that will be deployed, as well as the frequency of new instrument arrivals or displacements calls for the implementation of a system that will respond dynamically and autonomously to configuration changes. Moreover, the vast amount of data produced (petabytes will be available in the archive after few years of operation) signals the need for powerful, efficient and intelligent data processing and analysis systems that will mine the live as well as archived data streams to detect trends, classify content and extract features, feeding the results back into the master database, thereby turning raw data into information. Finally the information will be transformed into knowledge by the scientists. Both of the above aspects are challenging, but approaches based on recent cyber-infrastructure concepts will allow the use of innovative solutions to the problems at hand.

The proponents of this project believe that the deployment of a Service-Oriented Architecture relying on Web Services will provide an elegant solution to the first problem above, whereas workflow orchestration techniques will be instrumental in helping scientists assemble complex processing chains to be executed amidst an ubiquitous grid infrastructure. While not solving the actual scientific problems related to data features discovery, this project's aim is to empower NEPTUNE and VENUS users to conveniently weave their algorithms and data sets into a data and processing fabric.

The data volumes from some underwater instruments -in particular if several days or weeks worth are requested at once- will be such that bringing them across various grid nodes will have to take place using fast computer networks, possibly through the use of User controlled light path (UCLP) techniques, as those have the ability to dynamically make bandwidth available on demand.

The present project will produce results in three key areas:

- Through the use of web services that will implement the communication with them, instruments will be integrated into the overall observatory cyber-infrastructure
- Really Simple Syndication (RSS) feed technologies will be deployed to deliver to event information to subscribers (scientists or processes)
- Workflow orchestration tools will be made available to facilitate the elaboration of user-driven, complex, scientific analysis applications that will be executed on grid compute resources

The objectives of the project will be evaluated at multiple points. Firstly, scientific oversight of the project will be guaranteed through the participation of a NEPTUNE and/or VENUS scientist who will participate in the tests preparation and execution, the project communication and the

¹VENUS and NEPTUNE Canada are presently funded by CFI/BCKDF (\$10.3M and \$62.4M, respectively) to install powered, electro-optic cables and many observatory nodes in coastal and deep-sea environments in the North-East Pacific Ocean respectively. They are pioneering new network technologies and the introduction of power and the Internet to quantitatively monitor and interpret all aspects of the marine environment in real time, which will transform the Ocean Sciences. Both projects are led by the University of Victoria (UVic). Both represent consortia, with NEPTUNE Canada leading 12 universities coast-to-coast. For more details about the projects please see <http://www.neptunecanada.ca/> and <http://www.venus.uvic.ca/>.

use cases preparation. Secondly, we are proposing an external peer-review of both the architectural and system designs at each of their respective delivery times. The success of the implementation and tests will be judged by way of a workshop that will convey a panel of VENUS and NEPTUNE users. (VENUS will have been fully operational for 12 months by the time this project is completed). LOOKING² personnel will review the results from an inter-observatory access point of view and will be invited to participate in the tests themselves.

2 LOOKING: The Laboratory for the Ocean Observatory Knowledge INtegration Grid. A complete definition of all acronyms used in this document can be found in **Appendix A**.

Participants and Their Contributions

The University of Victoria (UVic), which is hosting both the VENUS and NEPTUNE projects, will support the project by offering the use of its premises and infrastructure. UVic, a leading BC university, has over 18000 registered student and over 3000 staff and faculty. UVic commands a yearly cash flow of over a quarter billion dollars. NEPTUNE Canada is a consortium of 12 Canadian Universities. The UVic team will be composed of:

- Benoît Pirenne, Assistant Director, IT, for NEPTUNE and in charge of the DMAS development will act as an overall coordinator for this project.
- Barbara Liang is NEPTUNE Canada's Manager, Finance and Administration. Barbara will be primarily responsible for the day-to-day resource management of the project.
- Chris Barnes is Project Director of NEPTUNE Canada since 2001. He will assist in the oversight, co-ordination and applications of this CIIP work with other NEPTUNE researchers.
- Verena Tunnicliffe is Project Director of the VENUS project since 2001. She will assist in the oversight and coordination with VENUS researchers.

Besides the VENUS/NEPTUNE DMAS Project Team, we are proposing to collaborate closely with **IBM Canada Ltd**, Markham, Ontario. IBM Canada is a key contributor to the Canadian economy through significant R&D investment, job creation, use of Canadian suppliers and extensive participation in university research programs. IBM Canada is one of the country's largest R&D investors, contributing \$334 million dollars in 2004. Its export revenue for the same year was \$1.7 billion. At year end 2004, IBM Canada and its wholly-owned subsidiaries employed some 20,000 regular full-time and part-time people across the country. In addition, IBM provided temporary employment for 2,938 people including 662 students, and we hired 1,274 regular full-time employees. In a recent KPMG / Ipsos-Reid Survey, IBM Canada was ranked among Canada's Top 25 most respected companies. IBM Global Services is the largest information technology services provider in Canada.

For the present project, IBM can contribute a very significant array of expertise to supplement the NEPTUNE and VENUS teams knowledge in the field of Web Services, Service Oriented Architecture etc. through its Hursley, UK and Rochester, MN Research Laboratories as well as its Pacific Development Centre in Vancouver. IBM will apply their cumulative knowledge in the fields of Earth and Life Sciences, petroleum exploration, manufacturing and finance, as well as healthcare to further this exploratory effort.

- Robert Heuchert will lead the IBM participation and will be key to allocating IBM's resources to the project and providing advice at all stages.

Please note that more details about the key participants are provided in Appendix B. Moreover, appendix H will contain a copy the agreement between IBM and UVic.

Other partners and related initiatives will be providing support and advice at various moments during the preparation and execution of the project. Although not formally part of the project, we expect occasional participation from:

- The Laboratory for the Ocean Observatory Knowledge INtegration Grid (**LOOKING project**), a US National Science Foundation-funded research effort into the identification, synthesis, and assemblage of existing and emerging concepts and technologies into a coherent viable cyber-infrastructure design. The goal of this effort is to federate ocean observatories into an integrated knowledge grid: (key personnel: Matthew Arrott). The expected contribution of LOOKING is in the area of overall cyber-infrastructure architectures, coordination with similar initiatives south of the border as well as in the evaluation of the project's progress.
- The **Scientific Workflow Automation Technologies Laboratory**, San Diego Supercomputer Center, UCSD: (key personnel: Ilkay Altintas). The pioneering work in the area of workflows and grid interfaces will be essential to the workflow aspects of this project. In particular, the efforts with the Kepler toolkit are expected to be of high relevance for the present enterprise.
- The **Monterey Bay Aquarium Research Institute (MBARI)**, Monterey, CA. (Key personnel: John Graybeal, Luis Bermudez). The work presently carried out at MBARI in the area of ontologies (eg., the Marine Metadata Interoperability project (MMI)) will be instrumental to the success of this work in what concerns its interoperability with other international initiative by providing advice on our ontology choices. Other collaborations with MBARI are expected in the area of streaming data analysis, which this institute has been pioneering (e.g., the work of Duane Edgington).
- The **GridX1 consortium**, a Canadian computational grid, has offered to host the heavy data analysis applications that we intend to deploy to demonstrate the instrument and archive link with the grid. All of the GridX1 resource centres are linked with CANARIE and make use of the Globus toolkit. Key personnel: Dr. Randall Sobie, UVic).
- Finally, the **ORAN** with which we will be interacting is **BCNET's**. The project has been discussed with Mike Hrybyk, President. BCNET will assist this project's team in the preparation of the data transport requirements both from the instruments and the related shore station (Sidney, BC and Port Alberni, BC) as well as from the University.

Functional placement of the coarse grained services above can be translated into a specific node deployment view or “Run-Time” model illustrated below.

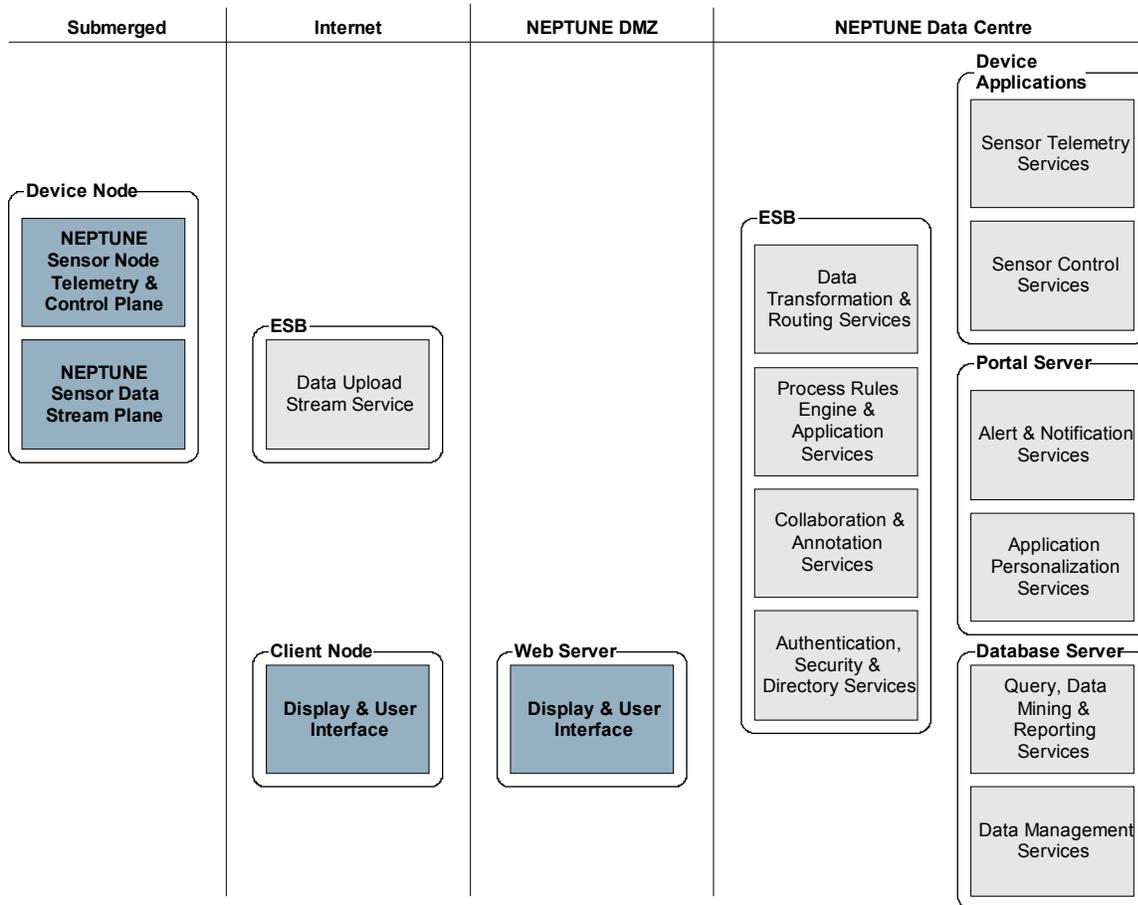


Illustration 2: Run-time view of the TSOA project services placement

Much, if not all, of the node and service interactions are mediated by the distributed Enterprise Service Bus (ESB).

In general an ESB should support:

- Request/response communication between loosely coupled SOA business components
- One-way message delivery for sending notifications to event-driven business components
- Reliable (store-and-forward) message delivery and publish-and-subscribe communication
- Some type of name space or directory which resolves a logical SOA service name to a specific software server implementation at run time
- Routing rules for topic-based, property-based or content-based message routing
- WSDL, SOAP, HTTP, XML, JMS, J2EE, Java APIs for XML-Remote Procedure Call [JAX-RPC], Java Business Integration [JBI] and other (non-XML) message formats and proprietary programming interfaces and messaging protocols

- Metadata for routing using XML Schema Definitions (XSDs) and WSDL files
- An intermediary is essential for ESB mediation to provide address indirection, rules-based routing (including publish-and-subscribe), message transformation and many quality-of-service (QOS) features
- Load balancing and failover
- Security
- Transformation
- Message validation, logging and auditing
- A run-time and development-time business services repository/registry which may include UDDI
- Integration via RPC, MOM or Web Services based technologies

Please note that a complete Services definition and description is provided in in Appendix C.

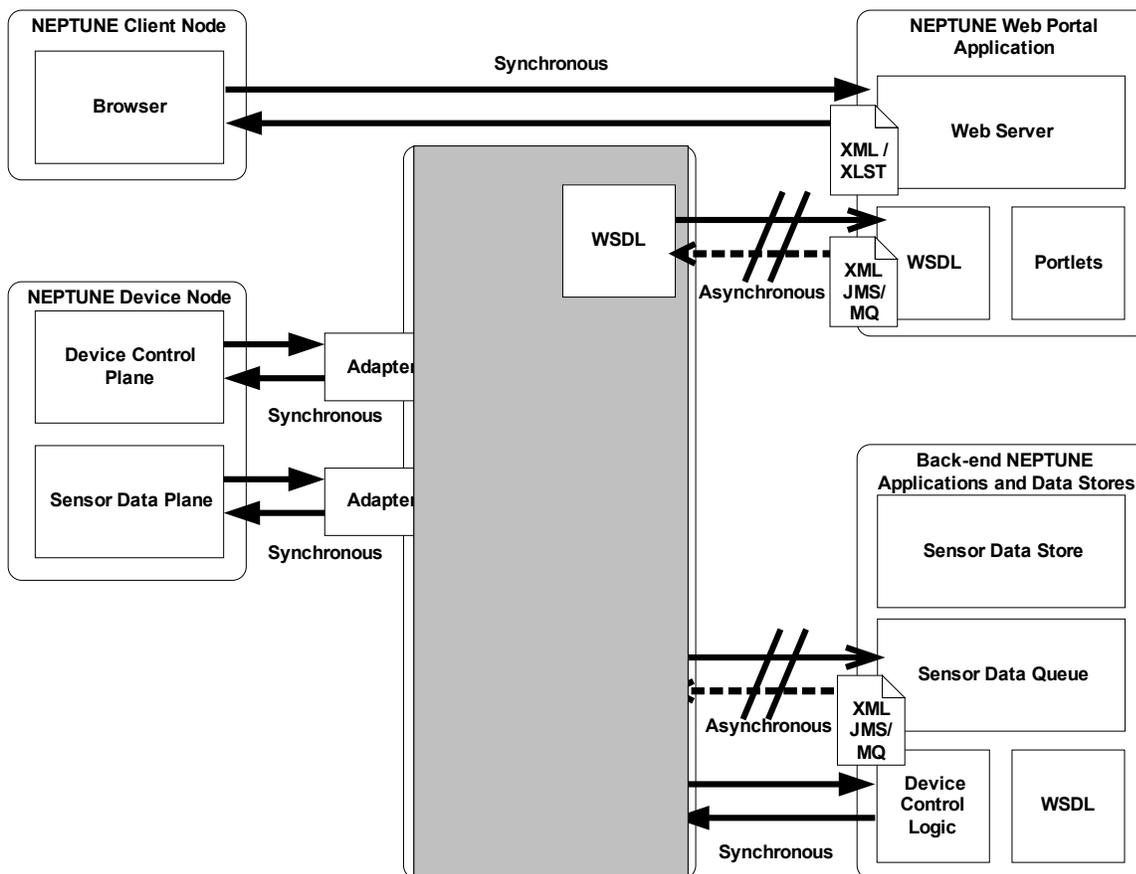


Illustration 3: Run-time view of the TSOA project Enterprise Service Bus mediated node interactions

2. Use case scenario example

Use case scenarios will be refined in the course of the project execution. So the following description will only serve as an illustration for what this project intends to accomplish. This single example will attempt to integrate the various concepts that the project is trying to address.

A marine biologist is studying large mammal populations off the West Coast. NEPTUNE is for her an opportunity to monitor the populations for several decades, if only at a few different locations. Her idea is as follows: as soon as a hydrophone records the characteristic signature of a large marine mammal, turn the camera in the direction of the sound source and record images for immediate viewing and later analysis.

In this plan, we can identify the following requirements on a system such as the one we are planning:

1. The hydrophone is recording continuously
2. Software is available to analyze the audio stream "live" and recognize the signature of the animals the scientist is looking for.
3. There is a system in place to warn the scientist of the detection
4. There is a facility in place to turn on a camera with no user intervention.

The way the case would unfold is as follows:

The scientist is provided with a tool to create **scientific workflows** that would have the following elements/capabilities:

- send real-time FFT audio sample to a spectrum analysis and classification system running somewhere on the **grid**. The system would have the extra capability to determine crude sound provenance and signal strength.
- an acoustic pattern matching system (on the grid) that would take as input the output of the spectrum analysis and classification system as well as a library of acoustic signatures identifying nearby marine mammals.
- a decision making feature that will connect to further action in case of a positive match.
- a automated scan/sweep sequencer that can invoke the camera system near the hydrophone to take pictures of the environment upon request. This would happen through the use of **exposed Web services**.
- the capability to relay images and video to our user, as well as the information about a positive detection using **RSS feeds**.

As can be seen in this example, the scientist is in a position to prepare and submit very elaborate event detection and reaction systems that will do the work on her behalf. The present project has allowed her to move into the era of "autonomous science".

3. Beyond project completion

The results from this particular project will be immediately usable in the practical environment of the VENUS observatory for which we will have the first deployment of the technologies developed here. In this respect, it is interesting to note that this will coincide with the deployment of the results of another CIIP project on the VENUS array: the HDTV proposal by John Roston et al. We will clearly have interactions with them, especially with the set up of UCLP from the observatory shore station to McGill University and the acquisition of data from their instrument. This will be done in a concerted way and the UCLP interfaces that we will use should be common between the two projects.

Beyond the initial deployment on VENUS, and using some the existing resources available in the NEPTUNE DMAS budget, we will be in a position to further the development of this project by:

- The integration of more and more instruments in the pool of supported ones for interaction over web services will be increased gradually.
- The addition of analysis applications supported through the workflow orchestration tool will be further improved so as to allow scientists to easily integrate their applications in a

grid processing environment and efficiently make use of the data available either on-line or from the archive.

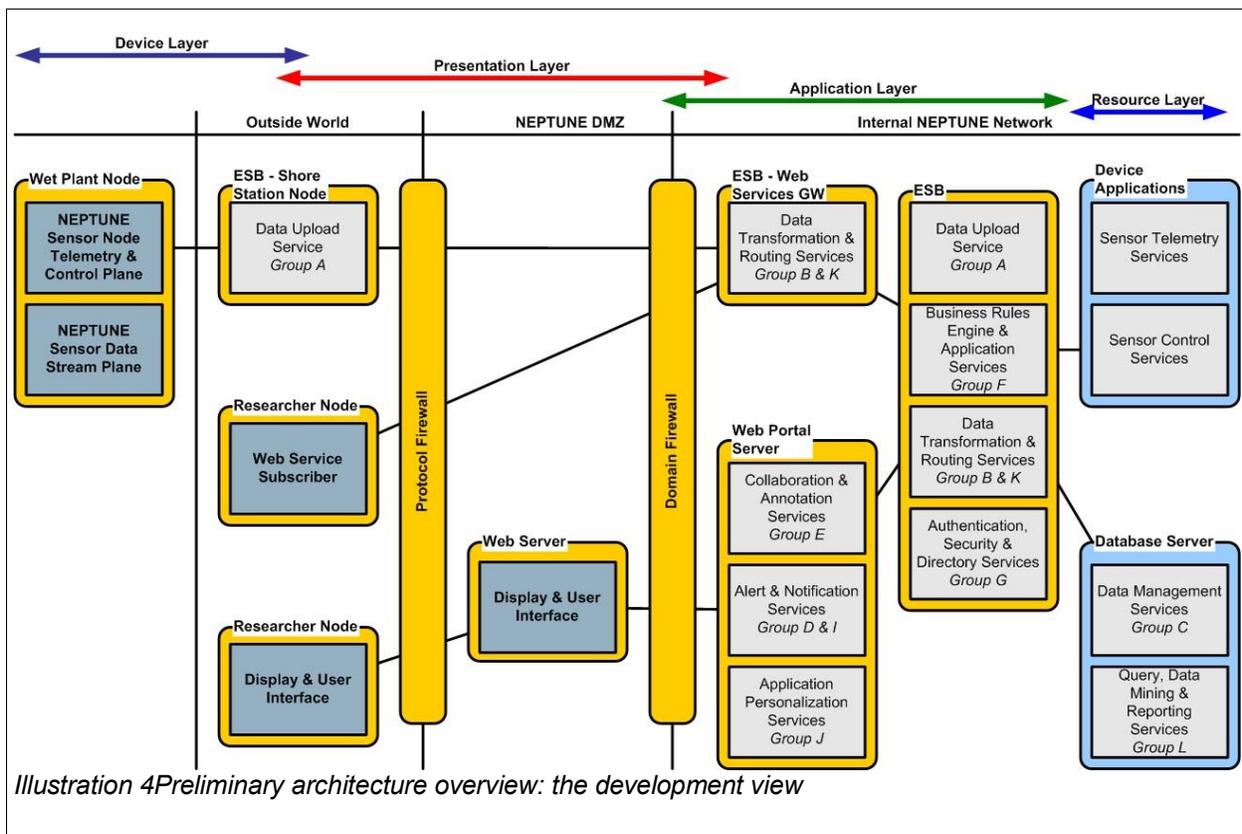
- Facilities to make selected data available to users will be further improved so as to allow the reporting of events as well as their the execution of the appropriate responses.

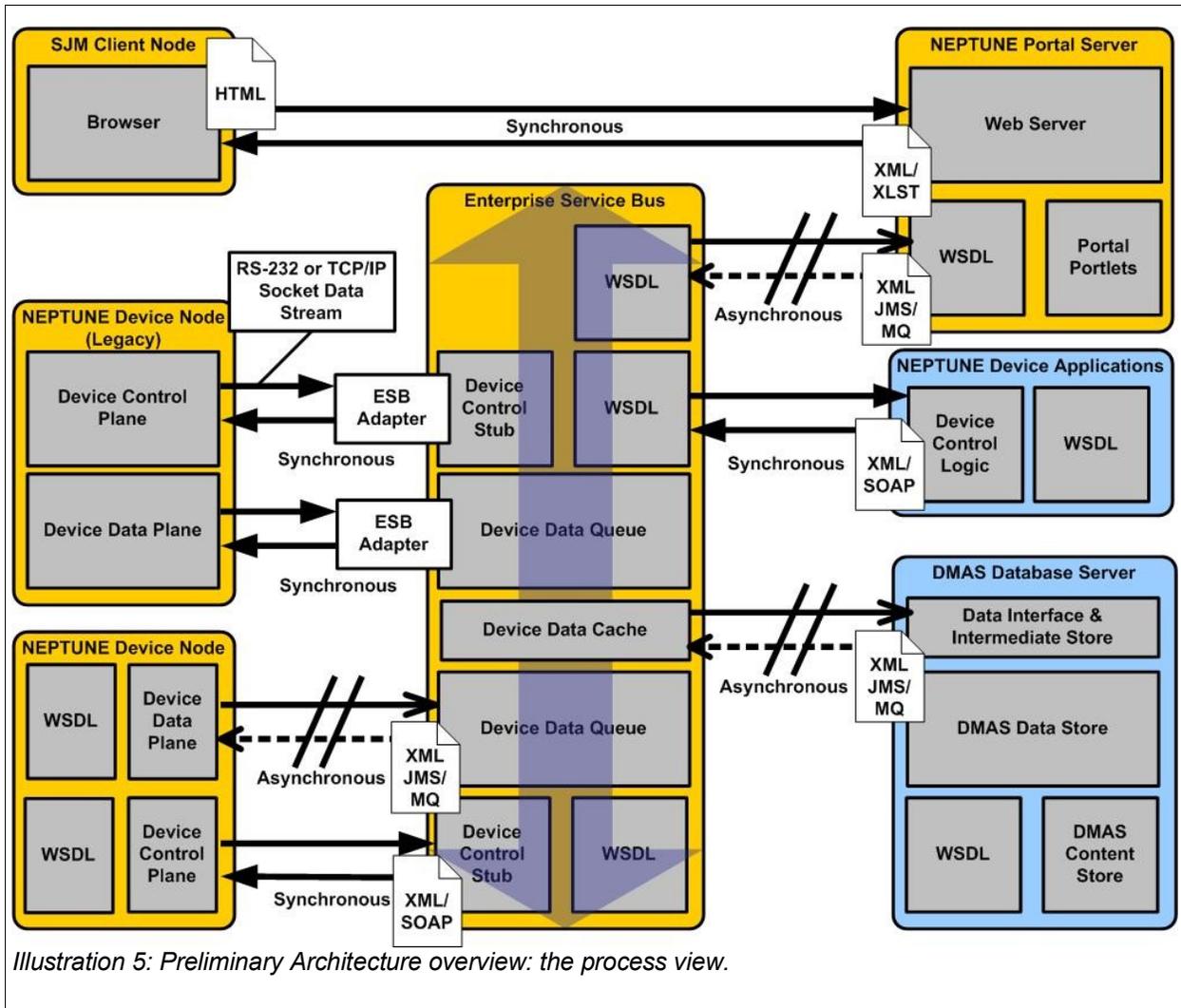
Thanks to CANARIE support, this work will allow NEPTUNE Canada to establish a leading position in the Big Science world, not just for the Observatory itself, but also for the science-enabling technologies that will have been developed.

4. Implementation details

Most of the implementation details will be refined in the project definition phase, however:

- Our involvement with IBM will allow the use of IBM's own Web services platform: Websphere MQ, which they will mostly contribute free of charge to this project.
- A preliminary architecture overview is depicted below both in a development view and as a process view.





- Security aspects are of course essential but are not addressed here as we are presently analyzing three options: password-based, token-based or Kerberos-style. We are proposing to deal with this issue during the detailed planning phase.

Project Selection Criteria

Discussion on the evaluation criteria.

1. Mandatory Criteria:

- a) This project is a **collaboration** between novel, groundbreaking ocean science projects (VENUS and NEPTUNE) led by the University of Victoria and IBM Canada Ltd. The former institution has a high potential for benefiting from new data management technologies, which the latter can clearly provide. While IBM Canada will place the focus on the basic technologies, independently of the application area, the other consulting partners we have listed before (LOOKING, MBARI, UCSD) will allow us to coordinate the efforts with other ocean-related initiatives (e.g., MARS). The current NEPTUNE and VENUS budgets do not allow for much spending in the software and data system area.
- b) This project will be **innovative** in that it will be the first regional ocean observatory to extend the Internet into the ocean. Further, the present planned use of Web Services technologies, of workflow tools to design ad hoc, complex, virtual data processing systems that will carry data over CANARIE's high bandwidth networks to remote grid processing centres is **unique** and so far limited to the high energy physics and astronomy's virtual observatory initiatives.
- c) Relying solely on the connectivity provided by CANARIE to link our shore stations to a BCNET transit-exchange (about 200km), this project's **use of CANARIE networks** will be quite extensive: beyond the shore station, the data centre located on UVic's premises will be extensively used by our scientific community to acquire archived or quasi on-line data for the purpose of processing/visualizing either at their own institution (one of 12 Canadian Universities) or at multiple grid compute centres such as those of GridX1 etc.
- d) This project is part of UVic's NEPTUNE Canada project. NEPTUNE Canada is funded to build and instrument an undersea network as well as to collect the data produced. The goal of this project is to quickly bring into the data management part a significant array of new technologies in a short time scale. Given the current budget situation, this would not have been possible. The CANARIE CIIP program will be instrumental in providing an unique opportunity to implement those technologies. The detailed project plan highlights what can be achieved in the available time frame. Beyond the project deadline, **sufficient funds from the initially guaranteed DMAS budgetary envelope will be available to ensure completion:** We indeed intend to retain some of the manpower hired for this project to pursue the work. Moreover, NEPTUNE Canada already has excellent human resources available for financial control and project management. The evaluation of the project outcome will be carried out at multiple times by tasking system experts to review plans, progress and accomplishments. Finally, the project outcome will be disseminated through various public presentation and by making the software available for free download through the project web site.
- e) Based on IBM's WebSphere MQ environment, the software elements developed in the framework of this project will have a strong long term support capability and should be easily portable to various similar environments. Moreover, with a planned lifetime of over 25 years for the Observatory, there will be a **long-term commitment** by the NEPTUNE Canada project team to maintain and further develop this product. NEPTUNE Canada is the first large scale cabled observatory. Others are being planned (ARENA in Japan, ESONET in Europe) and **might be interested in adapting our technologies.**

- f) The CIIP project discussed in this document will be **implemented in Victoria, BC**. UVic will be the home for the software developers hired for the project.

2. Desirable characteristics:

- a) The concept we are trying to develop here is primarily one where an **SOA** based on **Web Services** concepts and involving **workflow tools** is in place to support the data management from the observatory.
- b) The applicants definitely have a strong experience in the use of web services, SAO and workflow concepts. **IBM**, the prime partner in this project **has significant resources world-wide in the field of web services** that are at the disposal of the project. Moreover, the external expertise that we are involving both in the preparation as well as during the execution of the project is world-class:
- Matthew Arrott from the NSF-funded LOOKING project will help and guide our architectural design and make it match the requirements of ocean observatories;
 - John Graybeal, Luis Bermudez, of the Marine Metadata Initiative (MMI) at MBARI will provide ontologies support;
 - Ilkay Altintas, lead developer on the UCSD Kepler project, will contribute to the workflow and grid access aspects of the present undertaking.

A first preparatory project workshop was organized on July 6-7 where all above-mentioned individuals and partner representatives were present.

- c) Web services will be made available throughout our data management system, both **at the interface with the instruments** as well as access point to both our measurements **database and file servers**. As many as possible will be exposed to users.
- d) IBM Canada and UVic are not planning to commercialize the products of this development. A GNU-like public-domain licence would be characterizing the outcome, allowing others to contribute freely to the existing baseline that we will provide.

Detailed Project Plan

1. Project Management

The project will be managed as an entity reporting both to UVic's NEPTUNE Canada as well as to IBM Canada Ltd. through this company's science and healthcare department.

The project itself will have a management component resting on the participation of the PI (the project director), the finance manager and of a project manager. Three focus groups will be created, each containing the necessary manpower (see WBS). External advisers will provide the necessary support to each group as well as to the management and logistic. Project oversight is provided by the VENUS and NEPTUNE Project Directors.

The roles and responsibilities of each participant is shown in Table 1 below.

<i>Participant/Consultant</i>	<i>Role</i>
UVic/NEPTUNE Canada DMAS	Project leadership
IBM Canada Limited	Principal project partner
MBARI/MMI	Provide ontology support
UCSD/Scientific Workflow Automation	Provide advice on workflow systems
LOOKING	Provide advice on Service Oriented architectures

Table 1: List of participating institutions (in bold) and consulting partners (regular) together with their respective project roles.

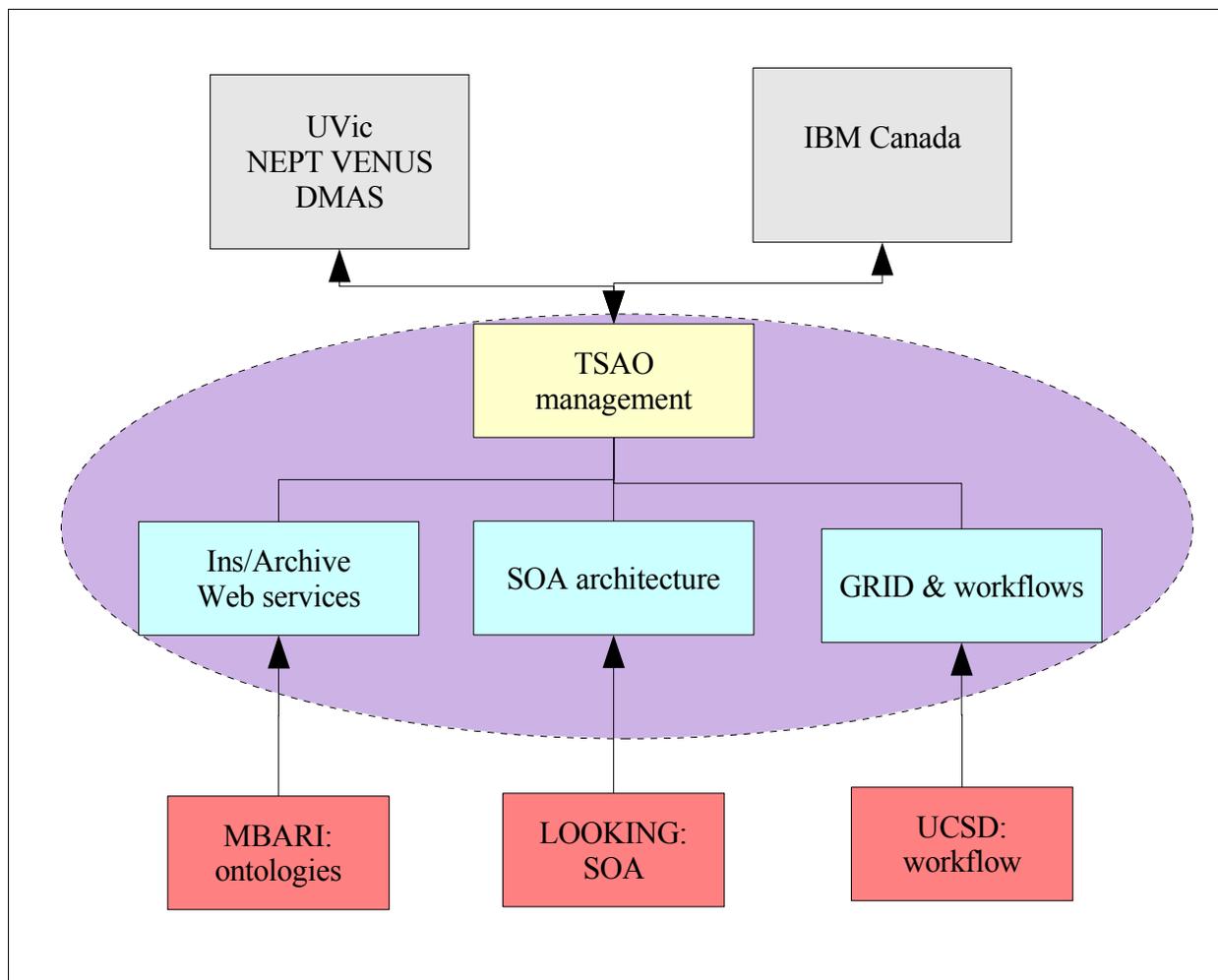


Illustration 6: Proposed project org chart: the purple oval denotes the elements of this project; the red boxes the external advisers and the grey boxes the parent organizations

Possible disputes within the project team will be resolved through the provisions to that effect contained in the UVic R&D agreement that will be signed with IBM Canada Ltd. This contract contains effective dispute resolution clauses. Differences will be resolved by arbitration. For smaller contractors and consultants, a "Business Firm Retainer" will be in place that will define the extent of our relationship with those external bodies. Where necessary, assistance will be provided by UVic's "Innovation and Development Corporation" (IDC).

Should a minor consultant withdraw from the project, the impact will be modest, as solutions for replacing his/her function would be easy to implement. Only slight delays might have to be suffered. In case IBM Canada would pull out, a larger issue would arise in that key expertise and software contribution might be in jeopardy. To mitigate the consequences of this unlikely event, the agreement with IBM will have a clause to that effect whereby the software provided will be kept by the project and any hardware received will be reimbursed. Human resources provided by IBM as part of their contribution to the project would have to be replaced, leading to inevitable delays in the final deliveries.

2. Project Plan

Introduction:

The project's main objective is to provide the VENUS/NEPTUNE DMAS with a Service Oriented Architecture that will enable sensors as well as retrospective data access, making use of CANARIE's CA*net 4 network and of UCLP technologies to transmit vast amounts of live or historical data to visualization systems or grid processing nodes.

Another key objective is the automation of sub-sea event management: scientists will, using workflow orchestration tools, assemble software agents that will autonomously perform the detection and pre-processing of events, set off alarms and trigger reactions. Those same workflow tools will be used to assemble access and processing components (including authentication, data retrieval, data transmission, data analysis programs, visualization systems) to define dynamic analysis infrastructures that will use high-speed networks and high performance computers on the grid.

Seamless interoperability with other emerging Ocean Observatories (e.g., in the US, Japan or Europe) linked through the concepts that LOOKING is advocating is also an objective of this project: Some of the milestones of the present initiative will match those of LOOKING. Of particular interest would be the ability to offer VENUS data for joint tests. Should the MBARI MARS project express interest in our initiative, we would clearly involve them as well. Given the time line of this project, we should be in a position to be perform joint tests with LOOKING's Ocean Observatory Data Services Platform by the time their development reaches version 2.0 for both real-time and retrospective data services.

Deliverables:

Given the short time frame for the development of this project, the proponents expect that the following items will be delivered (as reflected in the work breakdown structure):

- instrument integration into the overall observatory cyber-infrastructure through the use of web services: we will attempt at having two types of instruments connected in this way: a simple used to measure physical parameters of the environment as well as a more complex one requiring more sophisticated control. Moreover, web services will also be in place to perform archive queries and receive results.
- Using the outcome of the above implementations, RSS feed technologies will be deployed to deliver to event information to subscribers (scientists or processes)
- A limited set of Kepler primitives will be developed to allow the control of analysis applications to be executed on grid compute resources. They will clearly hide a lot of the complexity related to data access (including authentication, query and retrieval), grid resources access (authentication, job submission, results collection) as well as all that is required to transport the data across from the instrument or data centre to the grid resource available (e.g., UCLP).

This last set of items clearly rests on the availability of the other two as well as on the existence of readily available primitives to deal with GridX1 and UCLP.

Methodology:

The NEPTUNE Canada – IBM project team has elected to adopt the IBM Global Services Method for the realization of this project. The Method is a proven and mature approach used

extensively by IBM practitioners and customers. It is built around six fundamental best practices:

- Develop iteratively
- Manage requirements
- Use component-based architectures
- Model visually
- Manage change
- Continuously verify quality

The IBM Global Services Method provides detailed guidance on how to perform project activities. This guidance is provided as engagement models (templates) to IBM practitioners and is adopted and tailored for each project. These engagement models range from custom application development (e.g. Rational Unified Process, RUP) to implementation of Enterprise Resources Planning (ERP). They also can contain IBM sub-disciplines such as Object Oriented Application Design (OOAD) and Service Oriented Modelling & Architecture (SOMA)

The NEPTUNE & IBM project plan is based on the Solution Configuration (Lite) engagement model. It is an issue-based project approach which is used on medium sized projects that leverage commodity software (e.g. middleware, portal servers) as well as existing assets whose code and associated documentation are available (e.g. CANARIE, DMAS, Kepler). These assets are then encapsulated and facades are constructed to meet specific client's interface needs.

The defining characteristic of a package-like asset is that it is substantially a black-box component which is configured to meet particular requirements by setting parameters, by writing in a special purpose configuration language, or by writing extra subroutines to fit around the package. (In contrast a white-box asset would allow the developer to work directly with the design and code of the asset.)

The current NEPTUNE Canada – IBM project approach is based around work products in five domains:

- Research, covering a lightweight version of Ocean Sciences research transformation constrained by the capabilities of the asset
- Organization, covering change management, NEPTUNE organizational redesign and implementation, and education and training
- Application, covering the analysis, design, coding and testing of the adapted asset, the building of interfaces to sensors and other systems and the integration and system testing of the complete solution
- Architecture, covering the assessment of existing infrastructure, design of the new SOA-ESB infrastructure³ and the processes and procedures to run it
- Project Management, covering deployment and the those aspects of project management specific to solution consulting and services

Solution Configuration (Lite) uses a subset of Solution Configuration work products:

- There is less background work done leading up to the formulation of the business proposal.
- Organizational redesign is omitted. Change management is still included, though.

3 A SOA infrastructure map is shown in **Appendix D**.

- There is less analysis of the current infrastructure, and less work on the processes and procedures and on the capabilities of the IT department.

Solution Configuration has four phases:

- Proposal contains activities that are specific to projects of this kind which can be added to a more complete Proposal phase
- Assessment and Planning contains:
 - analysis of the current Ocean Sciences environment, and formulation of the proposed research solution
 - analysis and design of the proposed new infrastructure
 - analysis of the gaps between the requirements and the capabilities of the asset and their high level resolution
 - detailed planning of the rest of the project
- Design, Build and Test contains the configuration and test of the package, the detailed resolution of gaps between the requirements and asset capabilities, the development of interfaces as well as any custom interface development.
- Deploy and Rollout contains acceptance and full rollout of the solution.

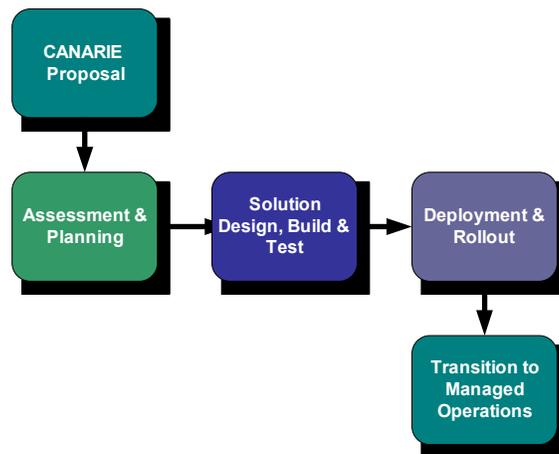


Illustration 6: NEPTUNE CIIP project overview, using the IBM method

NEPTUNE–IBM CANARIE – CIIP project approach addresses the initial three phases; “CANARIE Proposal”, “Assessment & Planning” and “Solution Design, Build & Test”

Overall the NEPTUNE–IBM CANARIE – CIIP project approach consists of a series of “Phases” and “Activities” which themselves consist of “Tasks”, “Sub-tasks”, “Input Work-products”, “Output Work-products” as well as “Perform Roles” and “Assist Roles”.

The following are the aforementioned project Phases and high level Activities:

CANARIE Proposal

- Define NEPTUNE Project
- Assess Asset / Application Fit and NEPTUNE Interface Needs
- Assess NEPTUNE Infrastructure Needs

Assessment and Planning

- Plan NEPTUNE Project
- Start NEPTUNE Project
- Initiate Assessment and Planning
- Define Research Solution
- Refine Solution Requirements

Infrastructure Design and Planning	Plan and Prepare for Testing
Plan Development	Infrastructure Implementation
Monitor Project	Interface Program Build and Test
Handle Exceptions	Package Configuration and Test
Confirm End of Phase	Integration and System Test
Solution Design, Build & Test	Monitor Project
Initiate Phase	Handle Exceptions
	Confirm End of Phase

The schedule of activities is shown in the WBS in Table 2. This schedule is based on an official project launch in early October 2005, assuming notification from CANARIE by mid-September. Please note that we will assume here that phase I (Proposal and Proof of Concept) has been done (and resulted in the present document), whereas the last phase (Deployment and Roll-out) will not take place as part of the present project.

This project plan has been defined at the occasion of a special-purpose preparatory technical workshop that took place on July 6-7 in Victoria. All partners and prospective advisers were present or represented. The refinement of this proposal is the result of the work performed at that meeting.

Phase, tasks and subtasks	Scheduled Finish	Estimated Cost
TSOA	5-Jan-07	978400
Assessment and Planning	15-Sep-06	244480
Project start	3-Oct-05	0
Finalize TSAO plans	18-Oct-05	13440
Project Mngt Plan, start staff	3-Nov-05	11520
Risk assessment	11-Nov-05	5760
Analyze existing environment	3-Feb-06	79200
Negotiate use of GridX1 resources for V/N	3-Feb-06	57600
Identify applicable grid technologies	9-Dec-05	14400
Other environment	23-Dec-05	7200
Change management (external communication)	3-Oct-05	960
Define research solution (internal communication)	28-Feb-06	16320
Usability & acceptance definition, test plan	13-Mar-06	8640
Adapt TSAO/DMAS architecture	30-Jan-06	11520
Document existing env & stds	24-Feb-06	13680
Architectural design	5-Apr-06	20160
Implementation plan	1-May-06	12960
Monitor/end of phase	2-May-06	0
Project web site prepared	15-Sep-06	21520
content preparation	17-Oct-05	9600
integration in environment	31-Oct-05	5600
set up and release	1-Nov-05	720
web site maintenance	15-Sep-06	5600
Project manager/SOA specialist hired	12/12/05	28800
Solution Design, Build and Test	5-Jan-07	733920
Initiate Phase	2-May-06	0
Detailed Organizational requirements	4. May. 2006	2880
Change management (ext. comm.)	5. May. 2006	960
Test plan ready	11. May. 2006	3600
Working environment ready	12. Oct. 2005	0
grid specialist hired	28. Feb. 2005	19200
Workflow specialist hired	28. Feb. 2005	19200
Web services programmers hired	25. Nov. 2005	38400
Staff training on Web Sphere technology	19. Jan. 2005	7200
Use cases revised	18. May. 2006	8800
Use cases revised (Workflow)	3. Mar. 2006	2400
Use cases revised (Ins access)	3. Mar. 2006	2800
Use cases revised (event feed)	18. May. 2006	3600
Analysis	26. May. 2006	19520
workflow and Grid applications	31. Mar. 2006	9600
ins access	17. Mar. 2006	5600
event feed	23. May. 2006	2160
review	26. May. 2006	2160
Programming Cycle	22. Dec. 2006	457600
Workflow and grid applications	8. Dec. 2006	144000

Phase, tasks and subtasks	Scheduled Finish	Estimated Cost
Develop required WS for grid work	26. May. 2006	28800
KEPLER primitives development	26. May. 2006	19200
Grid-enable two data analysis applications	17. Mar. 2006	28800
Integration of UCLP technology	18. Aug. 2006	28800
Maintenance and expansion of KEPLER primitives	8. Dec. 2006	38400
Instrument Access Development	22. Dec. 2006	188800
Develop instrument 1 WS with input from other projects	1. Sep. 2006	67200
Develop instrument 2 WS	10. Nov. 2006	57600
Instrument 1 access WS maintained and expanded	22. Dec. 2006	44800
Instrument 2 access WS maintenance	15. Dec. 2006	19200
Event Feed	8. Dec. 2006	124800
Develop Event detection/reaction capability	10. Nov. 2006	28800
Develop archive access (query & retrieval) WS	18. Aug. 2006	57600
Archive access WS maintained and extended	8. Dec. 2006	38400
User Support Material	22. Dec. 2006	20000
Review and adapt Kepler doc	9. Jun. 2006	4800
Produce event detection manual	24. Nov. 2006	4800
Produce instrument access manual	22. Dec. 2006	5600
Produce archive access manual	22. Dec. 2006	4800
Integration and System tests	22. Sep. 2006	10800
Prepare Integration environment	8. Nov. 2005	0
Integration tests performed	22. Sep. 2006	10800
System reviewed with LOOKING	2. May. 2006	0
Joint Review with LOOKING	8. Dec. 2006	0
Monitoring	11. Dec. 2006	115200
Final Report	24. Dec. 2006	9600

Table 2: WBS table showing the phases, tasks and subtasks identified in the framework of this project. The structure is based on, and adapted from, IBM's standard engagement model as described earlier. The costs indicated only represent a rough estimate of the direct manpower expenses incurred and do not include all of the overheads. Please refer to the financial schedules in the appendix for a precise account of the costs. Completed tasks are marked in highlighter colour where yellow denotes a completed activity and pink an on-going one. A line left in its original state denotes a task not yet started.

Intellectual Property

As suggested earlier in this document, the intention of the partners in this project is to make the code as well as the documentation developed in the course of this work available freely to any interested parties. A GNU-like public license will be suggested. The NEPTUNE Canada and IBM partners do not consider that they will be able to obtain revenue from this work, mostly of use to the scientific world at this stage. In this way, other projects will be invited to adopt the code and contribute to its further development.

Clearly, we will not redistribute directly code borrowed or purchased from third party, but rather simply make reference to it. This particularly applies to sensitive, export-restricted, security method that we might end up using.

Web Site Information

The web site describing the project will be available at the following address:
<http://arc1.neptune.uvic.ca/dmas/ciip/>

Appendices

Appendix A: Glossary

<i>Term/Acronym</i>	<i>Expansion</i>
ADCP	Acoustic Doppler Current Profiler
API	Application Programming Interface
BAM	Business Activity Monitoring
BCKDF	British Columbia Knowledge Development Fund
BCNET	British Columbia Academic and Research Network Operator
BPEL	Business Process Execution Language
CADC	Canadian Astronomy Data Centre
CFI	Canadian Fund for Innovation
CIIP	CANARIE Intelligent Infrastructure Program
DMAS	Data Management and Archiving System
DMZ	De-Militarized Zone
DND	Department of National Defence (in Canada)
ERP	Enterprise Resource Planning
ESB	Enterprise Service Bus
GNU	Gnu is Not Unix
GPS	Global Positioning System
HDTV	High-Definition TeleVision
HTTP	HyperText Transfer Protocol
IBM	International Business Machines
JMS	Java Message Service
J2EE	Java 2 platform, Enterprise Edition
LDAP	Lightweight Directory Access Protocol
LOOKING	Laboratory for the Ocean Observatory Knowledge INtegration Grid
MARS	Monterey Accelerated Research System
MBARI	Monterey Bay Aquarium Research Instiute
MMI	Marine Meta-data Interoperability
MOM	Message-Oriented Middleware
NEPTUNE	North-East Pacific Time-series Uundersea Networked Experiments
NSF	National Science Foundation (US)
OOAD	Object-Oriented Application Design
ORAN	Optical Regional Academic Network

<i>Term/Acronym</i>	<i>Expansion</i>
QoS	Quality of Service
RDBMS	Relational DataBase Management System
RPC	Remote Procedure Call
ROV	Remotely Operated Vehicle
RSS	Rich Site Summary (RSS 0.91) RDF Site Summary (RSS 0.9, 1.0 and 1.1) Really Simple Syndication (RSS 2.0)
SAML	Security Assertion Markup Language
SMTP	Simple Mail Transfer Protocol
SOA	Service-Oriented Architecture
SOAP	Simple Object Access Protocol
SOMA	Service-Oriented Modelling and Architecture
SQL	Structured Query Language
UCLP	User-Controlled Light Paths
UCSD	University of California, San Diego
UDDI	Universal Description, Discovery, and Integration
UI	User Interface
UVic	University of Victoria
VENUS	Victoria Experimental Network Under the Sea
WBS	Work Breakdown Structure
WS	Web Services
WSDL	Web Services Description Language
XML	eXtensible Markup Language
XSD	XML Schema Definition
XSLT	XML Style sheet Language Transformation

Appendix B: Resumes of key personnel

- Benoît Pirene, Assistant Director, IT, for NEPTUNE and in charge of the DMAS development will act as an overall coordinator for this project. He obtained a Master in Computer Science from the University of Namur, Belgium in 1986. Immediately following the degree, he worked for the European Southern Observatory (ESO) in Garching, Germany. In 2000, he was heading ESO's Operations Technical Support department. With a staff of 16, he was assuming the responsibility for a 30TB science archive and allowing over 3000 users to access ESO and Hubble astronomical data. During these years, he was a key participant in the development of ESO's end-to-end data system, recently awarded ComputerWorld's 21st Century Achievement Award recognizing world-class IT excellence. Benoît joined the NEPTUNE project in October 2004.
- Barbara Liang joined NEPTUNE Canada as Manager of Finance & Administration in August 2005. Before joining NEPTUNE, Barbara was Financial Controller for a private investment company. Prior to this, she held the position of Assistant Controller for 3 years with CellFor Inc., the world's leading independent supplier of high technology seeds to the global forest industry. Barbara worked with PRT Management Inc. for 2 years, where she was part of a team involved with the merger of two biotech companies, now known as CellFor Inc. At CellFor she established an administration department, which required the hire of personnel, development of policies & procedures, and the transition of the company's accounting system in-house from external resources. Barbara graduated from Capilano College in Business Administration and in March 2003 obtained her C.G.A designation. She has several years combined finance, not-for-profit, and private sector experience
- Chris Barnes is Project Director of NEPTUNE Canada since 2001. He will assist in the oversight, co-ordination and applications of this CIIP work with other NEPTUNE researchers. For the previous decade, he served as Director of both the Centre for Earth and Ocean Research and the School of Earth and Ocean Sciences at University of Victoria. Previous academic appointments were at the University of Waterloo (Chair, Earth Sciences (1975-81) and Memorial University (Chair, Earth Sciences, 1981-87) where established the Centre of Earth Resources Research. From 1987-89, he was Director General of the Sedimentary and Marine Branch, Geological Survey of Canada. He has managed a wide range of earth and ocean science projects. Barnes has served on many boards/councils and has authored or co-authored over 150 publications and 200 conference abstracts.
- Verena Tunnicliffe is Project Director of the VENUS project since 2001. She has worked for the past two decades in deep sea and coastal benthic systems exploring new ocean habitats and pioneering new approaches to deep ocean study. Her major tools for deep ocean work include submersibles, remote vehicles and deployed instruments. She has worked with engineering teams to develop observing techniques in the deepsea. Tunnicliffe led the first submersible expeditions to discover hydrothermal venting on the Juan de Fuca Ridge. She is also a contributor to the NEMO observatory on Axial Volcano - an array of battery-powered seafloor instruments placed offshore in the bathyal zone.
- Robert Kent Heuchert is a senior IT Architect with fourteen years experience in a diverse range of enterprise, integration and technology architectures. Currently with IBM Global Services, his project experience includes; Service Oriented Modeling & Architecture (SOMA), Enterprise Service Bus - Enterprise Application Integration (ESB-EAI), Web Services, high performance network computing technologies, UNIX and Linux clusters, high

availability LAN/WAN/SAN networking systems, storage technologies, directory services and enterprise systems management solutions. Serving in a leadership role on both internal and external solution development initiatives including such areas as application development (patent pending), business intelligence, customer relationship management, data/meta-data management, enterprise applications, informatics frameworks - infrastructures, Laboratory Information [Management] Systems (LIS/LIMS) and systems integration.

Appendix C: Service Definitions and Descriptions

NEPTUNE Sensor Node: the submerged NEPTUNE sensor device containing a diverse instrument package which will include some of the following instruments:

- Seismometer
- Hydrophone
- Magnetometer
- Electromagnetic Sensor
- Pressure Sensors
- Geodetic Monitors (Acoustic Transponders & GPS)
- HDTV
- Acoustic Doppler Current Profiler (ADCP)
- Borehole Thermistors & Seismometers
- CH₄, O₂ & SO₄ Sensors
- Profiler (winch drive device)
- Sonar
- Sediment Dynamics
- Still Camera – Web Cam
- GC/MS
- Temperature Sensors
- Crawlers and other remotely operated vehicles (ROV)

NEPTUNE Sensor Node Telemetry & Control Plane: with such a complex instrument payload contained in the Sensor Node, the Telemetry & Control Planes attempts to create a common internal bus for controlling the scientific instruments (the applicability of the IEEE1451 standard for sensors will be studied). Further the sensor employs the bus in order to communicate telemetry (health & safety) data for both instruments such as the ROVs as well as the Sensor Node itself. This is also the location of the Observatory resource management, where power and bandwidth allocations are decided according to safety and science priorities.

Display & User Interface: The Display and User Interface (UI) services are responsible for formatting web page views and content. This component takes as input the user profile, personalization information, collaboration settings and information retrieved from the connectivity services to display information to the user. The UI manifests itself both in a browser context which provides the interface for:

- Sensor data access, query or annotation
- Configuring personalization settings, alerts, thresholds or views
- Providing an RSS support for near real-time data stream monitoring
- Leverages the ESB for comprehensive message transformation facilities which applies XML Stylesheet (XSLT) transformation to render views

Data Upload Stream Service: A forward cache for capturing direct sensor output from the NEPTUNE node, temporarily caching this raw data, converting it into an XML message format (compatible for asynchronous message transfer) which is then streamed to the NEPTUNE Enterprise Server Bus (ESB) Data Transformation & Routing Services for appropriate rules-based conversion, manipulation and direction. The service provides:

- High performance and throughput characteristics (asynchronous communications)
- Strong quality of service capability where message flows are essentially transactions (using WebSphere MQ as a transaction manager)
- Use of WebSphere MQ as the Java Message Services (JMS) transport assures the once-only delivery of persistent messages
- High levels of availability will be achieved using multiple brokers and execution groups underpinned with WebSphere MQ clustering

Sensor Control Services: Utilizes a “rich client” for direct access to sensor instrument manipulation and interrogation via the NEPTUNE Sensor Node Telemetry & Control Plane. Although much of the control logic resides in the specific application it nevertheless relies heavily on the ESB for mediation, routing and message format transformation logic.

Sensor Telemetry Services: Employs a “rich client” interface for direct access to sensor health and safety or ROV/camera location/position information via the NEPTUNE Sensor Node Telemetry & Control Plane.

- Real-time establishment of instrument position or location in support of event data and Sensor Control Services
- Connectivity to instruments using WebSphere Business Integration Adapters

Data Transformation & Routing Services: combining the messaging infrastructure with message transformation and content-based routing in a layer of integration logic is the Data Transformation & Routing service which provides between service consumers and providers.

The main aim of this resource is to provide virtualization of the diverse resources, allowing the process logic of the researchers to be developed and managed independently of the infrastructure, network, and provision of those business services. Characteristics include:

- Dynamic routing using database or service lookup.
- Support for request-response, fire and forget as well as publish and subscribe.
- Asynchronous and synchronous delivery support (i.e. data stream vs. control)
- Protocol transformation support (e.g. HTTP to JMS), decoding, encoding and header validation of SOAP messages (including faults).
- Able to act as an intermediary between a service requester and service provider, with independence between the two.
- Message, Event and Service driven processing
- Transaction management support and assured once-only delivery of persistent WebSphere MQ messages.

Process Rules Engine & Application Services: as the primary enabler for Business Activity Monitoring (BAM) capabilities, the process rules engine provides the means for rich process coordination and orchestration. Conditional processing takes the form of data transformation, message enrichment, event triggers for alerts or queries as well as providing for human intervention.

A benefit of adopting XML as the internal messaging format of the ESB is ability to attain the concept of “operational awareness”. As each message is a complete self-contained document available for content inspection and routing. The ESB can also provide value-added services by sending alerts on flagged messages or critical values in a “push” fashion. Specific characteristics include:

- Message flow processing can be adapted according to business rules
- Ability to define logic, such as dynamic routing, based on content
- The implementation of a service as a message flow can be changed without affecting the service requester.

- A service provider's implementation can change without affecting its access from a workflow or message route.
- A workflow can handle a service request with no, partial, or complete SOAP (or message) validation and processing
- Broad support for relevant data formats as well as process entity support through XSDs

Collaboration & Annotation Services: Collaborative services enable Ocean Sciences research communities or teams to communicate and share information more effectively using the portal. These functions include:

- Shared process or workflow design and coordination
- Threaded discussions or links to Blogs (i.e. web logs)
- Exchange, capture and persistence of interpreted comments, annotations or observations

Query, Data Mining & Reporting Services: Those services will provide:

- Execution on an scheduled or as-needed basis or as part of process orchestration (e.g. BPEL)
- Support for federation and aggregation where in the processing of a single service request from a client can be fanned out into several requests to service providers and aggregating the results into a single response.

Alert & Notification Services: Event-based computing is complementary with SOA. Both paradigms benefit from loosely coupled services. As events occur — for example, a pending lab order - an ESB-based system can trigger actions by calling other services. Through services, direct access to the process rules engine enabling quick analysis to determine what actions are warranted.

- Alerts will be able to take the form of Instant Messaging, e-mail (SMTP)
- Takes advantage of RSS support to providing near real-time data stream monitoring

Application Personalization Services: The personalization function enables the user to personalize their portal experience to their own liking. It allows users to:

- Customize their display, event triggers or notification services
- Set content preferences so content is targeted to their interests
- Collaborative filtering and rules based personalization

Authentication, Security & Directory Services: Provides access controls and directory services in support of both user to system and system to system interchange. Role based security governs user access while token based mechanism facilitate Web Services or Portlets. Specific functions include:

- Authentication, authorization and non-repudiation

- Services that are created from message flows could be published in an external directory such as UDDI. A client could then use UDDI to discover the services
- Ability to enforce confidentiality if necessary (i.e. Department of National Defence, DND)
- Incorporate current and emerging security standards (e.g. LDAP, Kerberos, WS-Security)
- Authentication and authorization for access over JMS can be provided by WebSphere MQ infrastructure
- Authentication and authorization for access over HTTP will be provided via a internal or DMZ based HTTP server

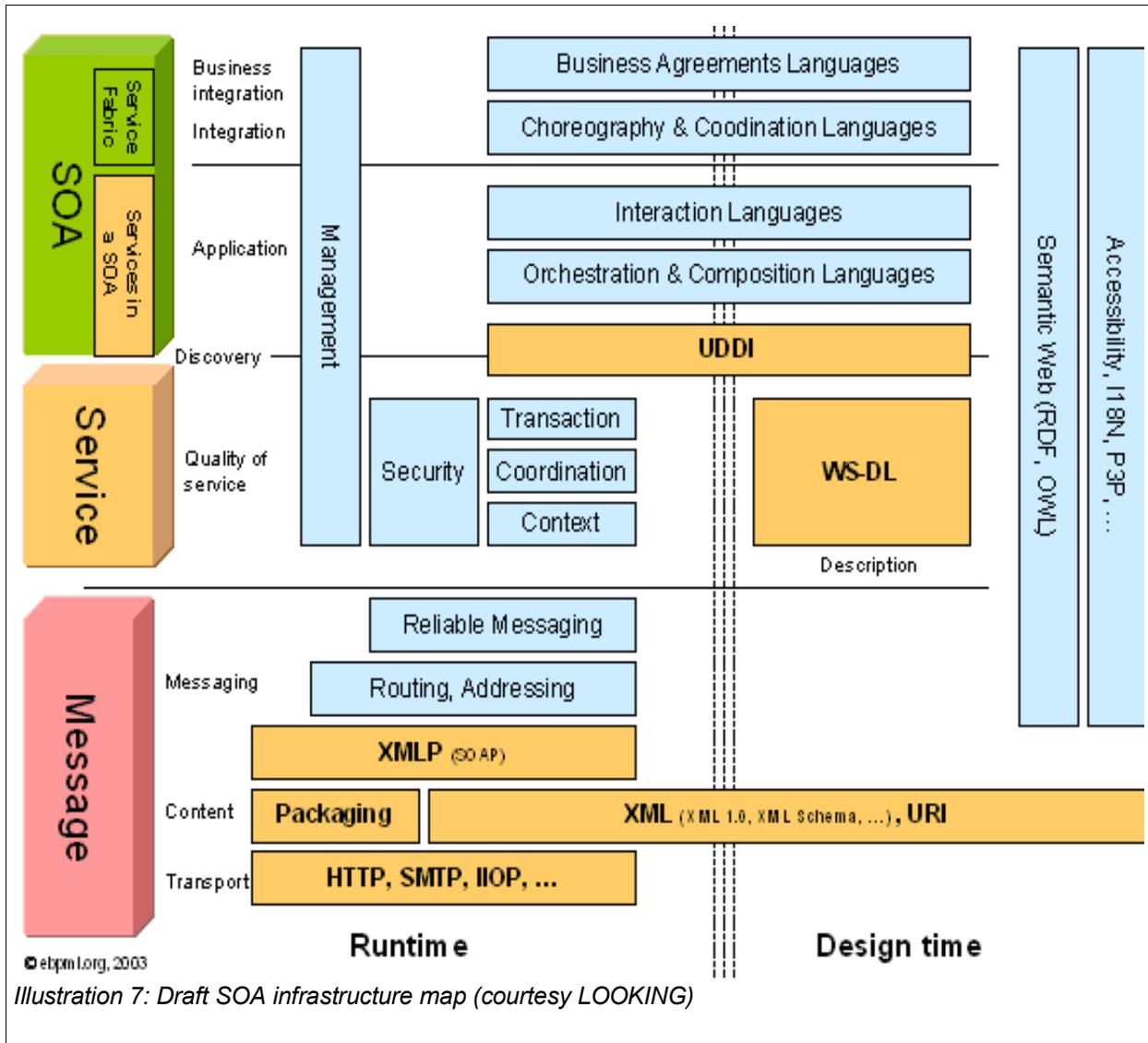
Data Management Services: Provide the ability to store both unstructured message and binary data in addition to normalized data structure for all sensor event, personalization, interpretive and transactional data. The normalized data structures are specifically built for retention temporal-based oceanographic data and metadata. The original raw sensor data is retained in XML or binary format and isolated from the relational data (for performance reasons) it is nevertheless logically linked back to the table spaces. Specific characteristics include:

- A robust and scalable Relational Database Management System (RDBMS) at it's core
- Observed and derived data are both persisted and archived
- Strong integration between relational databases and the ESB (which will be under transactional control) which bridges message based information (XML) with normalized data storage (SQL/RDBMS)

ESB will be used for data enrichment to provide additional information (data or metadata) that is required for service provider processing.

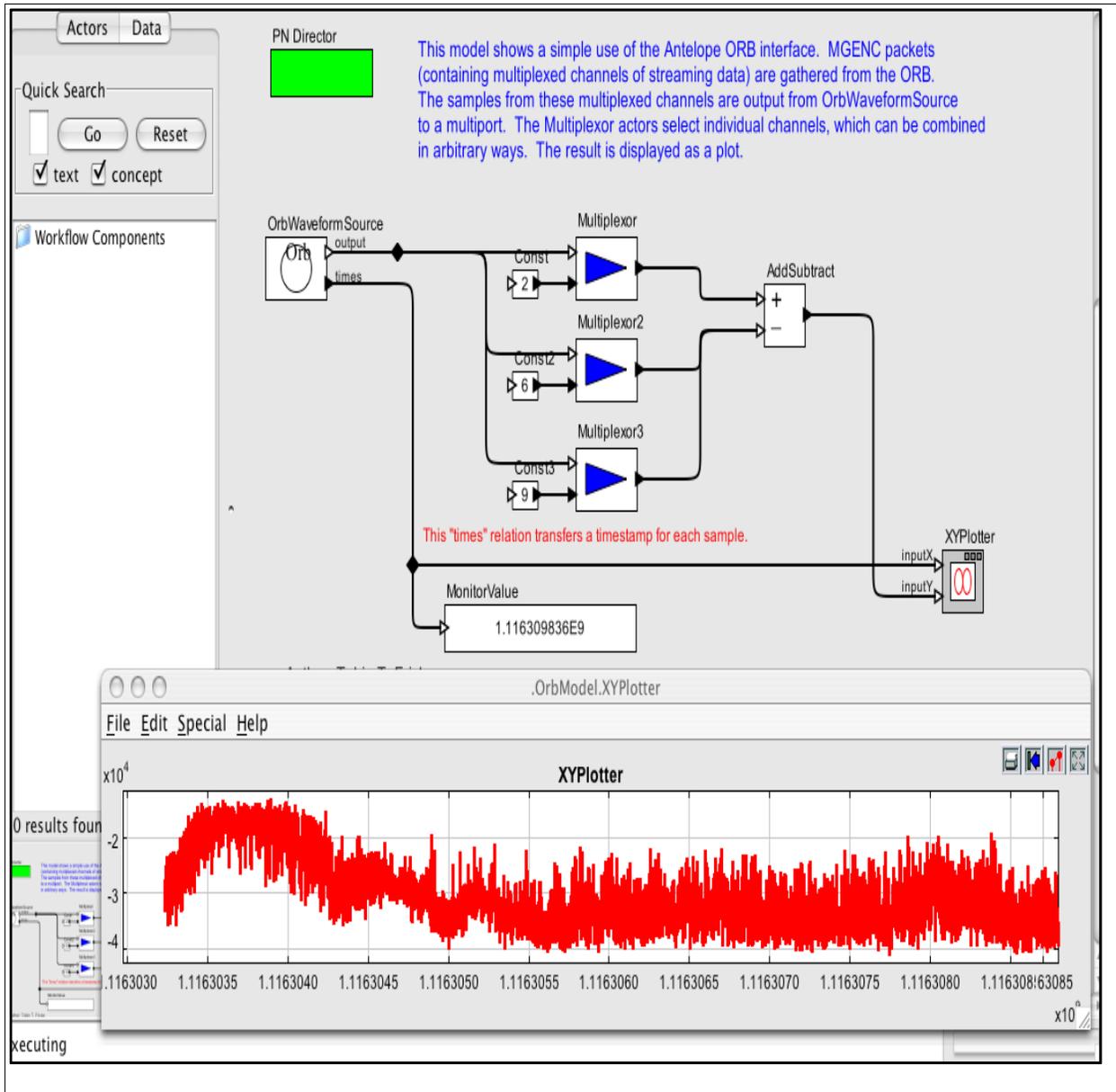
Appendix D: Draft SOA infrastructure map.

SOA infrastructure map (graph courtesy Matt Arrott, LOOKING project).



Appendix E: Screen shots

Example of a simple workflow preparation and execution GUI, which takes streaming input from an instrument and plots the output after optional processing.



Letter validating overhead costs



University
of Victoria

PO Box 3040 STN CSC
Victoria British Columbia V8W 3N7 Canada
Fax (250) 721-6221

Accounting Services

July 21, 2005

CANARIE Inc.,
110 O'Connor Street, 4th Floor,
Ottawa, ON
K1P 5M9

Dear Sir/Madam:

**RE: IIP-19 Toward a Service Oriented Architecture and Workflow
Management for VENUS and NEPTUNE
Confirmation of Overhead Rate**

As required, this letter will confirm that the University of Victoria is a member of the Association of Universities and Colleges of Canada (AUCC). The current agreement between AUCC, representing the universities, and Supply & Services Canada, representing the government, provides for a "one rate" system for overhead application to government contracts with Canadian universities. The "one campus" rate set by this agreement of 65% would be applicable to the University of Victoria's contribution to the above noted project.

Please do not hesitate to contact me should you require any further information or clarification.

Sincerely,

A handwritten signature in blue ink, appearing to read 'S. Bligh'.

Sandy Bligh, B.A.
Manager, Research Accounting
Telephone: (250) 472-4157 Fax: (250) 472-5196
E-mail: sbligh@uvic.ca

Appendix F: ORAN endorsement letter



July 22, 2005

CANARIE
110 rue O'Connor
Ottawa, ON K1P5M9

Re: Use of Regional Network for CANARIE Intelligent Infrastructure Program (CIIP)

To Whom It May Concern:

The University of Victoria and NEPTUNE are applying for funding for a project titled "Toward a Service Oriented Architecture and Workflow Management for VENUS and NEPTUNE." The proposed project will enable scientific instruments to be used with advanced software tools that utilize SOA. The project will have access to the VICTX in downtown Victoria and will make use of our regional network during this project.

BCNET enthusiastically supports this project moving forward.

Feel free to contact me if you have further questions.

Sincerely,

Michael Hrybyk
President/CEO

Cc: Benoît Pirenne, Morven Wilson, Chris Barnes

Appendix G: Project Budget

See the Excel Workbook.

Appendix H: Agreement with IBM

See the attached SoW of work, as co-signed by IBM and the University of Victoria and deliniating the extend of each partner's responsibility in the performance of the project.