Call for Proposals under the IMOS (EIF) Five Year Strategy: Enhancement or extension of IMOS – July 2009 to June 2013

Facility Project Plan template

Proposals should be submitted by 30 October 2009 to:
Tim Moltmann, IMOS Director, University of Tasmania
email: tim.moltmann@imos.org.au

Background:

This template has been provided to allow Facility and Sub-Facility Leaders, and other interested parties to prepare a Facility Project Plan following a call for proposals announced on 18 September 2009, with a closing date of 30 October 2009.

Prior to completing this template, please read the IMOS Five Year Strategy (the 'Strategy'), and Detailed Guidelines for Proposal Development (the 'Guidelines') – see the IMOS website at: http://imos.org.au/eif.html.

The Facility Project Plan must be in the following template and contain the information set out below:

Overview:

Proposed Infrastructure Investment:	Marine Water Analyses for Microbes at National Reference Stations
IMOS Facility:	Molecular Microbial Observatory
Operating Institution:	Australian Institute of Marine Science (AIMS) Sydney Institute of Marine Science (SIMS)/ University of New South Wales (UNSW) (SIMS/UNSW) CSIRO-Marine and Atmospheric Research (CMAR)
Facility Leader (for this Proposal):	Professor Linda Blackall, AIMS, 07 4753 4102, l.blackall@aims.gov.au
Other(s) key people involved:	Professor Peter Steinberg, SIMS/UNSW Dr Stan Robert, CMAR Professor Ricardo Cavicchioli, SIMS/UNSW Dr Torsten Thomas, SIMS/UNSW Dr Mark Brown, SIMS/UNSW Dr Federico Lauro, SIMS/UNSW Dr David Bourne, AIMS
Collaborating Institutions:	AIMS, SIMS, CMAR, UNSW

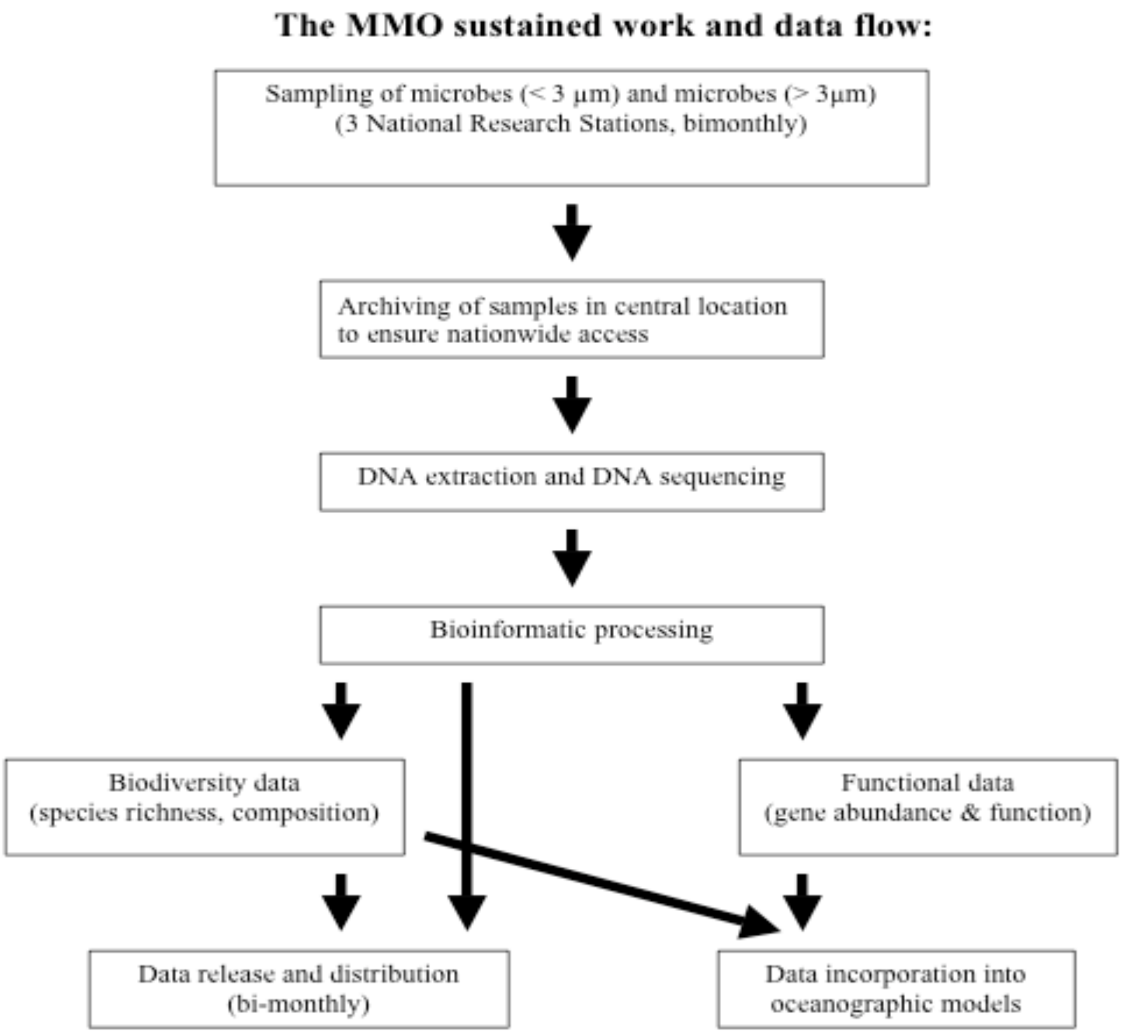
Please attach:

- All attachments will be emailed by Monday 2 Nov 2009.

Nature of Investment:

Implementation Strategy:
Summary. In the last 10-15 years, the capacity to study microbes in the world's oceans has been revolutionised by the development of molecular, genomic tools. This microbial "metagenomics" revolution has enabled the study of genes, species, communities and ecosystems in the marine environment in a way that has not been possible previously. The Molecular Microbial Observatory (MMO) will provide data on the types, abundance and activity of bacteria and archaea from three National Reference Stations (NRS) (Yongala, Queensland (Moreton Bay should be considered highly relevant and can be added); Port Hacking, NSW; and Maria Island, Tasmania; representing involvement of three IMOS Nodes). Water samples will be collected bimonthly and filtered, collected biomass will be archived, nucleic acids will be extracted and subjected to metagenomic sequence analyses. Sequence data representing an in-depth description of the microbial diversity and function of the water samples will be submitted to eMII so that they are available to all within the marine community. Integration with other IMOS Facilities and Nodes is a critical component of MMO since the interplay of microbial data with physicochemical oceanographic information from IMOS NRS's will lead to practical outcomes for oceanographers and other stakeholders in the marine biosphere.

Below is an outline of the MMO.



Objectives. The relevance of the marine microbial molecular revolution to oceanography is that marine microbes directly and indirectly control key parameters of ocean chemistry, including pH and CO_2 balance; light penetration depth, particularly in productive regions; and fisheries harvests in the coastal and open ocean.

Quote from Emeritis Professors of Marine Microbiology (Pomeroy, Williams, Azam and Hobbie, 2008):-

> We now know that every liter of "clear blue water" is teeming with a billion microbes—bacteria, viruses, and protists—far exceeding all multi-cellular metazoa in abundance, biomass, metabolic activity, and genetic and biochemical diversity.

The standard oceanographic measure of biological processes in the ocean is the easy to measure chlorophyll a, but this gives no information about species or abundance of microbes. Each individual microbial cell has an arsenal of metabolic capabilities such as carbon fixation, nitrogen fixation, denitrification and unique respiratory pathways that strongly influence (even control) ocean chemistry and productivity. A sufficiently detailed analysis of this capability and the dynamics of these process has not been possible because they take place inside microscopic cells and the measurements are time consuming and likely inaccurate.

The potential of the microbial molecular revolution is that it provides a pathway to identify, quantify and determine the function of the microbes present (most of which do not contain any chlorophyll a) in the marine biosphere. It offers the long-term promise of bringing biological oceanographic parameters to a level of accuracy and predictability that would make them robust components of general oceanographic models and in line with the accuracy and credibility of the comparatively on-line physicochemical data acquisition. What follows is a broad description of the revolution and how IMOS can play a critical, leading global role in this transformation.

The aim of this new IMOS facility (MMO) is to map biological observations and phenomena onto the physical oceanography already being conducted by IMOS. The underlying hypothesis is that changes in physical parameters – temperature, salinity, nutrients, currents – will drive changes in the key biological parameters targeted by a molecular observatory. We will provide data on the microbial biodiversity, abundance and function, while also contributing significantly to our understanding of ocean health and ecosystem functioning.

The backbone of MMO will be to take ongoing bimonthly water samples at three National Reference Stations from three IMOS Nodes (*building institutional strengths into a national capability*). These will be taken manually, as part of the existing NRS sampling regime. Microbial biomass will be isolated by filtration and archived at -80^{o}C. Nucleic acids will be extracted and also archived at -80^{o}C, and then processed by sequencing the entire microbial community DNA - called a "metagenomic" analysis. These sequence data can then be used to address a series of fundamental questions on microbial diversity and function, all of which can be related to the abiotic parameters generated from existing IMOS facilities.

A substantial number of additional samples will be brought into the IMOS data stream via co-investment. The NSW-IMOS Node have been doing metagenomics on marine samples for more than 3 years, funded from several sources. First, *via* a large ARC Linkage Grant (Kjelleberg, Steinberg, Thomas et al., $1.61 M, 2007-2011), metagenomics of selected coastal microbial communities from the Sydney region will be accessed. Second, past sequence and metagenomics data from the Southern Ocean will be made available by Cavicchioli, Lauro et al. This meets a key priority for IMOS - enhancing Southern Ocean observing. Cavicchioli and Lauro have recently received a 5 year ARC Discovery grant ($950, 000, 2010-2014) to continue this Southern Ocean sampling, and these data will also be made available to the IMOS MMO data stream. Combined, there are few if any such data sets in the world, and these data will be invaluable as comparisons to the NRS samples. Given the cost of sequencing these samples (mostly done in collaboration with major genomics institutes overseas), and the ship and expedition costs associated with the Southern Ocean sampling in particular, this co-investment has been conservatively calculated at *ca.* $3 M.

MMO will enable a characterisation of biodiversity at the levels of kingdom through to phyla, species, populations and genes, and then allow us to map this diversity, over time and space, onto the physical environment. For example, as the EAC ebbs and flows southwards, does it carry with it completely different biological communities? Similar to observations of the physical environment, this will provide us with the sustained observations necessary to track how the community changes in the context of climate change or other environmental drivers.

The data will provide substantial information about both *ecosystem functioning* and *ocean health (whole of system approach)*. Knowing community composition will also provide information about organisms with known functions (e.g., nitrogen fixers), as well as genes whose functions can be derived by comparison to existing genomic databases. This applies equally well to ocean health, where we can track in time and space both potentially dangerous organisms (e.g., toxic microalgae, ocean borne pathogens such as *Vibrio cholerae*) as well as actual virulence genes, which may or may not be associated with known pathogens.

Finally, we anticipate that due to rapid developments in DNA sequencing technologies that are occurring at unexpected rates, over the life of the project, the cost of this component could be halved concomitant with generation of a substantial amplification in data, thus driving down the cost per observation.

List of Major Activities.

Sample Collection

Staff from each NRS will be trained in the set-up and handling of the filtration system, through a 2-day workshop at the NSW-IMOS facility at SIMS. Filtration systems will then be installed on the boats that approach each NRS and customised on-board installation will require some cost for fittings, shelves etc. Two filter holders (one for the 3 µm and one for the 0.1 µm filter) will be required for each system and the suitable stainless-steel holder with a diameter of 30 cm (Millipore (@ $5000 per holder). In addition each filter systems needs two pumps (operating at >40 psi), hoses to connect holders and for water intake as well as some minor items such as buckets, tubes (for filter storages) and a 20 µm mesh. This will amount to an additional $1000 per filter system. The only consumable for the bi-monthly sampling constitute the filters (3 µm and 0.1 µm) of which 30, 36 and 30 will be required for year 1, 2 and 3, respectively.

Filter Archiving
Filters will be initially stored at -80°C at each IMOS node, but then shipped and stored at a central archiving facility – not requiring new freezers for the storage. Set-up of this facility requires the purchase of storage boxes and racks suitable for longer-term, low temperature exposure as well as barcoding/cataloguing with appropriate trackability. Filters will be shipped in bi-monthly intervals to the central archiving facility (at on IMOS Node, AIMS) from the other 2 NRS. Shipment on dry ice will cost $125 each.

Nucleic Acid Extraction and Archiving
DNA extraction will be done at the centralised archiving facility to ensure that consistent results for the extraction as obtained. Extracted DNAs will be archived at -80°C. A skilled technician can extract about 6 filters a week meaning that 180 filters will require about 10 weeks of work per year for 3 years. The annual salary for a technical assistant is $46,214 plus 30% on-cost, resulting in a cost of $11 553 pa. Molecular biology and chemical reagents are required for the DNA extractions at a cost of $5000.
This activity will start in July 2010 and will be completed in June 2013 (3 years).

Set-up of Bioinformatic Hardware
An experienced computational scientist is required to finalise the specification of the computational cluster as well as organising purchase and installation of the hardware. A 0.5 FTE equivalent to Academic Level A6 is requested to support this major task for 3 months. At a base annual salary of $68 483 plus 30% on-cost this amounts to $11 128.
Processing of raw DNA sequence data in the order of billions of nucleotides requires substantial high-performance computational capacity. Based on our previous experience, a cluster with at least 360 cores (equivalent to 90 nodes with quad-cores) would be required. We have recently built a cluster with 1/5 of the size (i.e. with 72 cores/18 nodes) for $60,000 and hence calculated that the HPC for this project will cost $300 000.
This activity will start in July 2010 and will be completed in December 2010 (3 months). This bioinformatics cluster will be housed at SIMS.

Set-up of Bioinformatic Software
An experienced computational scientist is required to set-up software and script that are necessary to process raw sequence data. A 0.5 FTE equivalent to Academic Level A6 is requested to support this major task for 3 months. At a base annual salary of $68 483 plus 30% on-cost this amounts to $11 128.
This activity will start in July 2010 and will be completed in September 2010 (3 months).

DNA sequencing
Extracted DNA will be shipped bi-monthly to the DNA sequencing facility (Ramaciotti Centre, see below) at a cost of $125 (must be done on dry ice) per shipment. DNA sequence data will be determined on the current state-of-the-art Illumina GSII sequencer. Total DNA of each sample will

be shotgun-sequenced in one channel of the machine. Cost per channel is quoted at $4110 by the Ramaciotti Centre at the University of New South Wales. In years 1, 2 and 3 we will process 5, 6 and 5 timepoints, respectively, resulting in 30, 36 and 30 samples (timepoint x 3-NRS x 2 size fraction). Once per year we will also amplify the small ribosomal subunit RNA genes for each DNA sample and sequence them in a multiplexing mode in one channel, resulting in one additional channel per year. Total cost for sequencing is therefore $127 410, $152 070 and $127 410 for years 1, 2 and 3, respectively.
This activity will start in July 2010 and will be completed in June 2013 (3 years).

Bioinformatic processing of data, Data distribution and public release and Incorporation of DNA data into oceanographic models
An experienced computational scientist is required to ensure processing, release and incorporation of the data. A 0.5 FTE equivalent to Academic Level A6 is requested to support this major task for 2.5 years of the project. A base annual salary is set at $68 483 plus 30% oncosts = $44 514 pa.
This activity will start in January 2011 and will be completed in June 2013 (3 years). Included here is the integration of additional data streams from coastal Sydney or the Southern Ocean as described earlier.

Major Parties Involved and Duration
Scientists listed above will be involved in most components of the data generation. Visits to the NRSs will coincide with regular IMOS trips, which are planned to occur monthly. There will be substantial opportunity for researchers to acquire bioinformatics skills from the team members at UNSW-IMOS and from the staff appointed on the project.

Major Equipment
The only major item is the high-performance computational cluster at $300 000.

Access, pricing regimes:
How will data access be provided?
Accurately annotated DNA sequence data will be delivered to eMII following extensive discussions with this facility. Storage of the annotated data does not require high computer processing capacity since the data will have been analysed to a stage that is practically useable by the marine oceanographic community.

How will data and products be managed?
The employed bioinformaticist in collaboration with Facility experts will set up a data handling and archiving system and will be responsible for most of the data management in collaboration with staff at eMII. Additionally, transparent communication to the Facility Leader will clarify the management of data.

What are the dependencies on external / other facilities (national and international)?
The sequencing of the DNA will occur at the Ramaciotti Centre at UNSW and consequently we are dependent on that facility for this task. We are also dependent upon the regular IMOS NRS monitoring regime, which is subjected to oceanic weather conditions at NRS's at the time of sampling. Sending of samples via the post (on dry ice) requires that we should be able to depend on this external service.

Collaborative structures for allocation of priorities
The team engaged in this project have been in close communication for a very long time, ranging from years to several months. Over this time period a collegiate feeling has been strong and although there is a Facility Leader, decisions will be made collaboratively.

Governance

Performance indicators

1. Establishment of the MMO Team. Documentation delivery.
2. Allocation of responsibilities to Team members. Documented.
3. Documentation of sampling strategy from NRS's, acquisition of first samples, reporting on first sampling, modification of sampling process from first sampling. Delivery of a detailed report.
4. On-going comprehensive reporting on bi-monthly sampling of NRS's and fine-tuning of sample processing depending on microbial molecular outcomes. Delivery of reports.
5. Determination of microbial metagenomic data in a timely manner from bi-monthly samples taken from NRSs. Billions of nucleotides in databases.
6. Bioinformatic analyses of microbial metagenomic data – annotation. Submission of annotated data to eMII.
7. Access to data by all in the Oceanographic cohort via eMII following extensive discussion between MMO and eMII.

Describe key risks and risk management strategies.

1. Collapse of Ramaciotti Centre, responsible for DNA sequence generation. This is highly unlikely and additionally there are numerous other facilities (in Australia and abroad) that can deliver these data, so the risk is quite low
2. Inability to access samples from the NRS's. This is likely to be a minor issue as alternatives to accessing the water samples from the NRS's will be quickly mobilised by IMOS. These would include additional boats to the moorings and changes in dates for data acquisition.
3. Loss of personnel. The proposed Facility is distributed across three IMOS nodes, all of whom have substantial expertise and capabilities in the relevant areas. Loss of any one or even several people can be readily accommodated by the distributed structure of the proposal.
4. Loss of stored samples. Replicates of all samples will be kept, both at the Node, which collected them, and at the central archive (AIMS).

- **Budget:** See attached Spreadsheet.

Facility budget proforma for IMOS EIF Call for Proposals 2010-2013

Enhancement to Facility / New Facility (*Note 1*)

		2010/11	2011/12	2012/13	Total
NCRIS/EIF Funding		(EIF)	(EIF)	(EIF)	(EIF)
Capital	High-performance computational cluster ($300,000), sample filtration apparatus modelled on Venter Global Ocean Survey (+$25,000x3 NRS)	380,000.00			380,000.00
Salaries	Technician ($11,553 pa); Bioinformaticist ($22,256 - setup, yr 1 only); 0.5 FTE Bioinformaticist ($44,514 pa - half yr in 10/11, full yr in 11/12 and 12/13)	56,066.00	58,309.68	60,552.36	174,928.04
Operating	Filters, Sample transport, Molecular Biology kits/reagents, Disposable Labware, Nucleic Acid Sequencing	137,410.00	163,070.00	139,410.00	439,890.00
	NCRIS/EIF Funding Total	**573,476.00**	**221,379.68**	**199,962.36**	**994,818.04**

Cash Co-investment		**2010/11**	**2011/12**	**2012/13**	**Total**
SIMS/UNSW	Bioinformatics hardware/software ($125,000); 0.4 FTE postdoc bioinformatics ($40,000 pa for 2 yr) ; Confirmed	165,000.00	41,600.00	0.00	206,600.00
	Cash Co-investment Total	**165,000.00**	**41,600.00**	**0.00**	**206,600.00**

In-kind Co-investment		**2010/11**	**2011/12**	**2012/13**	**Total**
AIMS	0.15 FTE of Linda Blackall Facility Leader, 0.05 FTE of David Bourne Project Collaborator (incl 1.5x multiplier); Confirmed	38,517.30	40,057.99	41,598.68	120,173.98
SIMS/UNSW	0.05 FTE of Peter Steinberg, Ricardo Cavicchioli, Federico Lauro, 0.1 FTE Torsten Thomas, 0.2 FTE Mark Brown - Project Collaborators (1.5x multiplier); Confirmed	62,000.00	64,480.00	66,960.00	193,440.00
SIMS/UNSW	Provision of data from ARC-Linkage Grant (Steinberg/Thomas - Venter Institute as Industry Partner) (80K over 3 yr); Confirmed	26,666.67	26,666.67	26,666.67	80,000.00
SIMS/UNSW	Provision of data from Southern Ocean metagenomics studies (Cavicchioli/Lauro - Aust Antarctic Div/ARC-D) (over 3 yr); Confirmed	1,755,000.00	500,500.00	500,500.00	2,756,000.00
CMAR	0.1 FTE of Stan Robert, 0.1 FTE of Guy Abell Project Collaborators (incl multiplier); Confirmed	48,970.04	51,418.54	53,989.47	154,378.04
	In-kind Co-investment Total	**1,931,154.00**	**683,123.20**	**689,714.82**	**3,303,992.02**
	TOTAL Resources	**2,669,630.00**	**946,102.88**	**889,677.18**	**4,505,410.06**

Acronyms	
AIMS	Australian Institute of Marine Science
SIMS/UNSW	Sydney Institute of Marine Science/University of NSW
CMAR	CSIRO-Marine and Atmospheric Research
FTE	Full Time Equivalent
ARC	Australian Research Council