Project: National eResearch Collaboration Tools and Resources

Project #: 2179

Contents:

Issue date: 20th September 2011

Responses must be received by NeCTAR by: 4:00 pm AEST Wednesday 02nd November 2011

ABN 84 002 705 224
The University of Melbourne
Parkville, Victoria 3010

NeCTAR is supported by the Australian Government (the Commonwealth) through the Super Science Initiative and the Education Investment Fund (EIF).

Table of Contents

Section 1 RFP

1.1 RFP Contact Details

RFP Proposals ONLY	proposals-rfp-nectar@unimelb.edu.au
RFP Questions ONLY	questions-rfp-nectar@unimelb.edu.au
General Queries **Questions relating to the RFP should ONLY be delivered via the appropriate email addresses above.**	The NeCTAR Directorate Room 3.11, Level 3 Doug McDonell Building The University of Melbourne, Vic 3010 Contact: (03) 8344 1277

1.2 RFP Timeline

The full timeline is published and maintained on the NeCTAR website at (http://www.nectar.org.au)

Request For Proposal issued	20th September 2011
Close for queries regarding proposal preparation	5 business days before the Closing Time
Responses to be received by (**Closing Time**)	04:00pm AEST 02nd November 2011

1.3 RFP Checklist

1. Have you registered online at http://www.nectar.org.au?	YES
2. Have you read and understood Part A?	YES
3. Have you read and understood the relevant project Part B documentation?	YES
4. Have you read and understood Part C?	YES
5. Have you completed all sections of Part D?	YES
• Section 2 Contact Information	
• Section 3 Compliance Statement and Departures	
• Section 4 Fields of Research (as appropriate)	
• Section 5 Response, noting the selection criteria in Section 6	
• Section 7 Milestones and Deliverables	
6. Have you asked any questions you needed to, and received sufficient answers?	
7. Have you returned the pack, Part D, to proposals-rfp-nectar@unimelb.edu.au?	

1.4 Submission Instructions

Proposals shall be submitted:

- electronically (as per Section 1.1);
- in English;
- in a legible font and size (suggested minimum 10pt);

- in text-searchable PDF; and
- a 10 Megabyte (including attachments) limit.

On Closing Time, the University of Melbourne Tender Board will issue all submissions to NeCTAR for review.

Proposals must be received no later than Closing Time specified in the Timeline or they will be treated as a Late Submission as described below.

NeCTAR reserves the right to change the Closing Time for any reason, in which event written notice of the change will be provided.

1.5 Late Submission

Proposals lodged after the Closing Time or lodged at a location or in a manner that is contrary to that specified in this RFP will be disqualified from the selection process and will be ineligible for consideration, except where the Proposer can clearly demonstrate, to the reasonable satisfaction of NeCTAR, that late lodgement of the Proposal:

a) resulted from the mishandling of the Proposal by NeCTAR; or
b) was hindered by a major incident and the integrity of the selection process will not be compromised by accepting a Proposal after the Closing Time.

The determination of NeCTAR as to the actual time that a Proposal is lodged is final. Subject to paragraphs (a) and (b) above, all Proposals lodged after the Closing Time will be recorded by NeCTAR, and will only be opened for the purposes of identifying a business name and address of the Proposer. NeCTAR will inform a Proposer whose Proposal was lodged after the Closing Time of its ineligibility for consideration. All such Proposals will be returned at the conclusion of the Selection Process.

Section 2 Contact Details of the Proposer

2.1 Proposer Contacts

The Contact Details of the Proposer are to be detailed in section 2.1.1 below.
Please add the details of any anticipated participating organisations in section 2.1.2. Add extra lines as required.

2.1.1 Proposer

Organisation Name	Monash University
Contact Name	Steve Androulakis
Position	Research Data Consultant
Business Address	Wellington Rd, Clayton, VIC 3800
Postal Address	Wellington Rd, Clayton, VIC 3800
Telephone	03 9902 9318
Facsimile	03 9902 9318
Mobile Phone	0401 493 730
E-mail	steve.androulakis@monash.edu

2.1.2 Participating Organisations

Organisation / Group Name	Location	Role
Monash University	Clayton, Vic.	eResearch Partner, Service Provider
RMIT	Melbourne, Vic.	eResearch Partner

2.1.3 NeCTAR Program

Please Indicate the NeCTAR Program to which this Proposal is being submitted (one only):

(eResearch Tool, Virtual Laboratories, National Servers Program, Research Cloud)

eResearch Tool

2.1.4 eResearch Tools submitted in support of a Virtual Laboratory Proposal

eResearch Tool proposals may be considered for funding under a special allocation, where they are submitted in support of a Virtual Laboratory proposal.

If this is a Virtual Laboratory proposal, list any separate eResearch Tool proposals which are being submitted in support of this proposal (please list Title and submitting organisation).

Not Applicable

If this is an eResearch Tool proposal submitted in support of separate Virtual Laboratory proposal(s), list those Virtual Laboratory proposals (please list Title and submitting organisation).

Characterisation Virtual Lab (Monash University)

If you are submitting an eResearch Tool proposal, and the Virtual Laboratory proposal is not funded by NeCTAR, indicate if you are requesting that the proposal be considered for funding as a stand-alone proposal.

☐ **No**

☒ **Yes**

Section 3 Compliance Statement

3.1 Proposed Sub-Contract Compliance

Are there any Departures from the Contract (Part C) Terms and Conditions?

☐ **No** There are no departures from the terms and conditions (i.e. Full Compliance)

☒ **Yes** There are departures from the terms and conditions

Detail the departures in Section 3.4 of this document.

The proposing organisation warrants that except for the departures listed in Section 3.4, the response is in full compliance with the Contract terms and conditions and no further contractual issues will be entered in to.

Signature of authorised person making the statement Name and role (printed)
Date

3.2 RFP Compliance

Are there any Departures from the RFP Terms and Conditions (Part A)?

☒ **No** There are no departures from the terms and conditions (i.e. Full Compliance)

☐ **Yes** There are departures from the terms and conditions (i.e. Does not Fully Comply)

Detail the departures in Section 3.4 of this document.

The proposing organisation warrants that except for the departures listed in Section 3.4, the response is in full compliance with the RFP terms and conditions.

Signature of authorised person making the statement Name and role (printed)
Date

3.3 Conflict of Interest

Are there any known or potential conflicts of interest responding to the RFP and its Terms and Conditions or in delivering the proposed works?

☒ **No** There are no conflicts of interest

☐ **Yes** Describe the conflicts in Section 3.5 of this document.

Do you commit to inform the University of Melbourne of any future conflicts or potential conflicts as they arise?

☒ **Yes**

Signature of authorised person making the statement Name and role (printed)
Date

3.4 Statement of Departures

	Clause or Reference	Nature of Compliance	Proposed wording of amendment
Proposed Sub-Contract			
RFP Terms and Conditions	1.1	Partially Comply	Monash asks for further discussion/negotiation in relation to two definitions under this clause. Specifically, Monash requests that the definitions of "Liability" and "Loss" be amended as follows: *Loss* includes any direct costs, expenses, losses, damages, charges or liabilities. *Liability* means any obligation, debt, fine, penalty, Loss or compensation of any kind.
RFP Terms and Conditions	4.2(d)	Partially Comply	Monash asks for further discussion/negotiation on this clause. Specifically, Monash requests that subclauses (ii) and (iii)

			be deleted.

3.5 Conflict of Interest

The Nature of the Conflict	Implications of the Conflict	How the Conflict is to be managed?

Section 4 Fields of Research

For RC and NSP proposals, this section is optional. RT and VL proposals must complete this section. Select up to five disciplines using either the two or four digit codes, or a mixture of both, and allocate a percentage score or weight against how closely the Proposal is aligned to a particular community discipline or Field Of Research. The FOR codes are available at: http://www.arc.gov.au/applicants/codes.htm

DISCIPLINE/FOR Code	Weight (percent)
02 – Physical Sciences	20%
03 – Chemical Sciences	30%
06 – Biological Sciences	50%
	100%

Section 5 Response Template

Complete the following table, ensuring a response to all headings and statements are provided. Attach this as a separate document to your proposal. Not all elements apply to every program (RT, VL, RC, NSP), please indicate "not applicable" as appropriate.

Any additional material or brochures can be added as attachments to the Proposal. The page counts are an indicative guideline for responses.

No	Title	Items to Address	Suggested Page Count
SUMMARY			
1	**Program and Proposal Title**	• *Which program is this proposal addressing? (RT, VL, RC, NSP).* • *Provide a short title for the proposal to use as a reference in communications.*	0.1

Bioscience Data Platform: TARDIS in the Cloud

2	**Executive Summary**	*Summarise the context that leads to this project and briefly outline the vision for the outcomes.*	1

Some of the most important challenges in biology over the next 30 years pivot on understanding the structure and function of proteins and their assemblies, due to their crucial role in the functions of organelles and cells. A structural characterization of the complexity of life on the cellular level, and accurate functional models of the biophysical and biochemical properties of large molecules such as those of proteins, DNA and RNA underpin modern medicine. This requires a multidisciplinary approach that integrates experimental techniques with computational methods. Monash University has a significant program of research focusing on characterising large multi-protein complexes; the research program at RMIT focuses on characterising the structure and function of biomedically important nanmoaterials. . This proposal seeks to build the computational resources that will be critical in attacking significant problems in biology, health and disease.

The Bioscience Data Platform (BDP) aims to bring existing computational systems together in a way that allows scientists to seamlessly work with data from capture through to publication. The platform will be backed by the Australian Characterisation Environment Virtual Lab to exploit the advantage of its uniquely powerful computational infrastructure. The platform will reach all areas of the structural biology workflow, from the inception of a research project through to scientific publication.

Typically, a scientist will collect data at an instrument facility (such as the Australian Synchrotron) to find their data catalogued and easily accessible on the storage cloud. The BDP

will provide a set of tools that link to high throughput computing infrastructure, potentially saving the scientist days of computational time and manual effort.

Visualisations will be available alongside text, and peers will be able to annotate and draw attention to all aspects, including text and visual elements. These same tools will assist the peer review process by enabling researchers to share data with scientific journals and eventually the public in a secure and rich environment that tells the full story behind a discovery.

The significant investment by ANDS for Data Capture and Metadata Store solutions will be leveraged to give scientists the ability to register these interactive data-driven publications hosted within the BDP to Research Data Australia.

The Bioscience Data Platform will:

- Offer the MyTARDIS data management system as a national service running on cloud storage hardware with linkages to major instrument facilities and HPC resources.
- Allow Australian research institutes and labs to register for the central web-based service and subscribe to feeds of their data and metadata from instruments such as those with existing MyTARDIS capability at the Australian Synchrotron and ANSTO.
- Enable Australian researchers to share data and collaborate via MyTARDIS by the implementation of the Australian Access Federation as a means of authentication and authorisation.
- Empower the biosciences community to conveniently leverage the power of high performance computing for complex simulations and analysis by developing web interfaces for commonly used research tools and linking them to HPC resources.
- Foster collaboration between research groups via rich data publications. Data annotation and interactive visualisation will provide the capability for researchers to highlight and discuss points of scientific interest.
- Provide a set of tools within the platform that works with data to generate information needed for submission to scientific journals and online databases such as the Protein Data Bank.

3	**Research Community Profile**	*Provide a profile of the research community that is sponsoring the proposal; include the aims of the community, geographic spread or location and membership size.*	1

The key participants and major sponsors for the BDP are Monash University and RMIT.

The research community sponsoring this proposal are associated with the structural biology research drivers and capability requirements of Monash University's Characterisation Virtual Lab. We will be building tools for two facets of research, linked by the need for high throughput data modelling and visualisation in the biosciences.

Structural Biology:

The structural biology community's research involves a wide range of biochemical, biophysical, and computational techniques and as such generates large and diverse datasets. The

understanding of protein structure now has a firm role in the molecular basis of all diseases, and as such is a vital underpinning for the future promise of drug design.

X-ray crystallography is the most common technique for the structure elucidation of proteins. For example, notable recent examples of its power and impact are the landmark structure determinations of the Ribosome, RNA polymerase II, a potassium channel, and F1 ATPase (Nobel prizes 2009, 2006, 2003 and 1997 respectively). Due to the enormous size and complexity of current biological problems we are currently witnessing a rapid and large scale implementation of a wide range of approaches within structural biology. This presents challenges for data management, analysis and publication. This is a truly global effort, involving tens of thousands of researchers.

Functional Nanomaterials for Biomedical Applications:

This community conducts research on the formation of functional nanomaterials using advanced crystallography characterisation, computational modelling and visualisation. The aim of this research is to understand the processes that create materials with enhanced physical, chemical and biological properties, typically in thin film surfaces. Such surfaces are to be delivered in new medical, biomedical and pharmaceutical materials and products. They may also have applications in energy in environment.

4	**Development Organisation Profile**	*Provide details of the organisations that will contribute to the development of the proposed infrastructure; include information about their capacity and capabilities, their track record and relevant experience in this role, their approach to quality standards, support and warranty mechanisms, etc. Include any supporting statements from Research Users. Where specific projects are mentioned, indicate specifically the aspects which support your track record or experience.*	2

The Monash eResearch Centre (MeRC) is a University-funded centre mandated to support and nurture eResearch. The Monash eResearch Centre was the first University eResearch Centre in Australia, established in 2005. It captures and combines Monash University's expertise in eResearch and existing technical infrastructure to develop Monash University as a leading international institution for modern research practice.

This research-led centre has at present 25 long-term appointed eResearch professionals, many of whom have research backgrounds, and is a blended unit of staff from various foundations of the university. The centre functions as a delivery vehicle for expertise and services drawn from the Office of the DVC-R, Monash's IT service division (eSolutions), Faculty of IT (incl MeSSaGE Lab), and the Monash University Library. The leadership resides with the research agenda of the Office of the DVC-R whereas the delivery is performed through the operational IT 'engine room' of eSolutions.

The centre's capabilities and expertise covers computational science, bioinformatics, software development, systems administration, visualization, data management, storage systems, agile project management and software engineering, and high performance computing. The

Monash eResearch Centre offers a complete integrated view of eResearch, from Research lab through to eResearch fabric, and has staff at all points of this continuum.

eSolutions leads and directs the provision of IT solutions to Monash University. eSolutions is currently leading substantial technological, service and organisational reform towards creating a single IT function for the University that operates according to the following vision - "partner with our customers to provide complete solutions enabling the Monash academic mission and delivery of the strategic agenda". They aim to be a customer focused organisation delivering, "flexible, responsive, coherent ICT services". eSolutions is the accumulation of 500 IT services staff across the Monash campuses.

This NeCTAR proposal is both developed and will be operated, by Monash University through MeRC and eSolutions. MeRC will form the researcher engagement functions and eSolutions will form the IT services portion of the service solution.

Because of MeRC's unique position within the university as a Research Technology Platform for the broader community it currently provides direct eResearch capability to around 3000 researchers around Australia. MeRC has successfully delivered capability into over 250 eResearch projects since 2007, and many of its eResearch tools and capabilities are used nationally (for example MASSIVE and Monash eResearch Collaboration Suites) and internationally (for example, Underworld, myTARDIS and NIMROD).

Monash University through MeRC and eSolutions have accumulated several eResearch "firsts" for Australia. Monash has pioneered and operated LaRDS, the Monash Sun Grid (MSG), VeRA and MASSIVE. LaRDS is the 4PB institutional data store dedicated to research, which at this scale is an Australian institution first. The MSG is the institutional heterogeneous grid of clusters, specifically commissioned to nurture researchers onto state and national-scale HPC infrastructure such as NCI and provide significant high-throughput computing capability. MSG was the first national infrastructure to include 'nodes' of very large memory 256GB and then 1TB beginning two years ago specifically enabling de novo assembly from next (2nd) generation gene sequencers of larger portions of the human genome. The three-year-old VeRA cloud was a national leading infrastructure in the provisioning of virtual machines to researchers. The VeRA service predates robust open source cloud software, but represents 150+ web-services oriented machines rapidly provisioned to researchers. MASSIVE is a specialised peak computational facility for the visualisation and analysis of primarily 3D characterisation data. Monash coordinates the partnership that also includes the Australian Synchrotron, CSIRO and the Victorian Partnership of Advanced Computing.

Recently Monash University (MeRC/eSolutions) has deployed a pilot instance of OpenStack. Monash is also in the midst of accreditation for ISO27001/2 (information security management) for research oriented IT. In partnership with the University of Melbourne, Monash has responded as a node provider to RDSI.

MeRC's Science Director leads Monash's Research and Development Laboratory for eResearch (MeSSaGE lab) and through this eResearch laboratory Monash University has been an active foundation member of the world's largest eResearch consortium, the NSF-funded PRAGMA. Furthermore, Monash University has formal links and exchanges with global eResearch powerhouses CALIT2, the San Diego Supercomputing Centre, NCSA and the Oxford eResearch Centre.

Prof Louis Moresi (Professor of Geodynamics) writes “At a recent international conference I made the case to the geoscientists that we have two pieces of world-class infrastructure at Monash which they should consider tapping into. One is the Australian Synchrotron that is straightforward for them to understand. The second is the geoscience cyber-infrastructure provided by the Monash eResearch Centre consisting of both hardware and eResearch professionals. Each of these pieces is substantial in its own right, but, put together, this creates an enviable facility that matches any other group in the world.”

RMIT eResearch: RMIT is active in eResearch through partnerships such as VPAC (established at RMIT and providing eResearch services to RMIT researchers) since the nineties, and has created an eResearch Office in 2008 with fully seconded ITS staff for HPC and information systems support. Additionally, there’s an eResearch software group predominantly supporting RMIT’s top CAT1 research grant performers and 4 strategically founded cross-organisational research institutes in Health Innovation, Platform Technologies, Global Cities and Design. The RMIT eResearch group also conducts projects under ANDS contracts and in collaboration with VeRSI under a recent partnership agreement.

RMIT's eResearch support includes a large visualisation cave, 3D virtual room and various visualisation walls, motion tracking and user interaction gear, mostly used for sports science, games, design and HCI studies in stress and emergency training. In 2012 with the opening of the multi-million dollar Design Hub, these facilities will be refreshed and consolidated in the Design Hub. This provides opportunities for multi-disciplinary engagements of the proposed project with advanced 3D visualisation capabilities in those other disciplines.

In both ANDS and VeRSI collaborations, variants of the MyTARDIS platform have been developed. The collaboration with Monash University is well established, around the MyTARDIS projects and beyond. It is expected that MyTARDIS skills raised in these projects will be available for the proposed project.

RMIT’s strategic research plan to 2015 includes a successively increasing allocation of resources to eResearch to support the university’s research growth. This strategic resourcing will be flanked by increasing ITS investment to guarantee sustainability of eResearch pilot projects through to production-quality delivery of research services into the future and beyond the sunset of ANDS and NeCTAR funding.

5	**Operational Organisation Profile**	*Provide details of the organisations that will contribute to the development of the proposed infrastructure once they have been commissioned; include information about their capacity and capabilities, their track record and relevant experience in this role, their approach to quality standards, support and warranty mechanisms, etc. Include any supporting statements from Research Users.* *Where specific projects are mentioned, indicate specifically the aspects which support your track record or experience.*	2

Development and operations is the same organisation – please refer to section 4.

The eSolution / eResearch partnership presently operates the following infrastructure for the Monash (South East) precincts -

- MASSIVE - a 1008 core with 160 GPGPUs cluster, approximately one year into operations. (The operation of this facility includes the Victorian Partnership for Advanced Computing.)

- LaRDS - Monash's four year old 4PB hierarchical data store dedicated for research purposes.

- Monash Sun Grid - Monash's collection of high-performance and high-throughput free (to affiliated users) computational resources. Presently at 1500 cores.

The partnership outsources the institution's need for traditional HPC requiring low-latency IPC (inter-process communication), for purposes such as finite element modelling, to NCI.

6	**Other Participants**	*Name any other institutions or groups that will need to be involved in the project planning and execution and their roles.*	0.5

The BDP project will consist of software developers in the eResearch groups of Monash and RMIT and members of Ashley Buckle's and Salvy Russo's respective laboratories. These members will be directly involved as advanced user-developers within the evolving platform.

While the project does not depend on other participants, workshops will be held with a broader community of potential users approximately half-yearly. This will ensure a participatory engagement with the broader stakeholder community and the preparation of uptake of the technology. The workshops will include all MyTARDIS users, in particular.

7	**Key Personnel**	*State any key individuals that are required for specific project activities and their availability.* *Provide names, organisational locations, and their expected roles.* *For example; Project Managers, designers, technical experts.*	0.5

Monash University:

Prof Paul Bonnington, Director - Monash eResearch Centre

A/Prof Ashley Buckle, NHMRC Senior Research Fellow – Faculty of Medicine

RMIT University:

Heinz Schmidt, eResearch Director – RMIT University

8	**Infrastructure**	*Describe the proposed infrastructure to be developed. Provide supporting documentation, specifications etc as required.*	2

The software infrastructure to integrate and host the Bioscience Data Platform's suite of tools will be developed by extending MyTARDIS.

MyTARDIS is a web-based data management system with a focus on the sharing, annotation and searching of large datasets.

An openly definable parameter model supports the ability for data to be described in technical terms for specific research domains. This flexibility in description has seen MyTARDIS used to hold raw data produced by instruments in areas such as macromolecular crystallography, small and wide angle x-ray scattering, infra-red micro-spectroscopy, x-ray photoelectron spectroscopy, neutron scattering and electron microscopy.

A standardised metadata ingestion service allows MyTARDIS to fit into an instrument's data capture workflow with minimal effort. Data can be made available on several levels of access, from single-user access only, to groups of users, temporary access via tokens or completely open access to all.

Instances of MyTARDIS can be federated, by configuring installations to subscribe to data from others installed data-producing instruments. A central index can be used to search for public data hosted across multiple deployments in a single action.

MyTARDIS 'in the cloud':

The Bioscience Data Platform (BDP) will be established as a nationally available, central service, by extending the MyTARDIS codebase to function on the NeCTAR Research Cloud and the integration of several widely used research tools.

Australian research groups will be able to sign up to the BDP and subscribe to feeds of data from instruments at MyTARDIS-enabled facilities such as the Australian Synchrotron and ANSTO's Bragg Institute. A web service interface will be available that allows new pathways for data from instruments and facilities that previously weren't MyTARDIS-enabled, to the BDP.

In addition to data subscription capability, users will be able to upload their own research data to MyTARDIS, in order to preserve, share and take advantage of the annotation and visualisation capabilities of the BDP.

Extending MyTARDIS into the NeCTAR Research Cloud will involve building support into MyTARDIS' codebase for NeCTAR's OpenStack middleware. Integration between MyTARDIS and the underlying storage will be all-encompassing, affecting the downloading, uploading and transfer of data into the system. Individual access to, and the sharing of data will be facilitated by developing an authentication plugin that implements the Australian Access Federation.

A Platform to HPC Resources:

An API will be developed for MyTARDIS that provides a software linkage to HPC resources. This will be developed to function with several existing HPC facilities such as MASSIVE, giving those signed up to the central MyTARDIS service the ability to seamlessly take advantage of these resources.

Specific scientific applications will be linked to this API in such areas as:

Molecular simulation:

Molecular simulation is a widespread technique used to study the dynamic properties of molecules, and generates significantly more data than a typical X-ray crystallography experiment. Typically, these studies use national computational resources such as NCI, VLSCI and MASSIVE, as well as local Monash clusters (estimated >5000 cores used by the Buckle group alone at any one time). There is an increasing trend towards merging and using this data with experimental biomedical data, presenting specific data management, analysis and publication challenges.

Computational Biophysics:

Salvy Russo's group has developed a tool and various scripts to manage different Monte Carlo simulations for biophysical models. These methods include Reverse Monte Carlo and Quantum Monte Carlo methods, for example. The latter method is the most accurate ab initio method known to produce highest quality calculations of the binding energy between various DNA/RNA bases and water molecules to date. This interaction is one of the most fundamental aspects of the structural behaviour of the molecular components of DNA/RNA and its accurate prediction is critical to the understanding of the interactions between large biomolecules such as those studied by Ashley Buckle's group.

The integration of such tools and their enabling and management on the NeCTAR Research Cloud and/or HPC centres supporting suitable licenses will be a core part of this project at RMIT.

RMIT's eResearch group has developed dcTARDIS, a variant of MyTARDIS for curating data collections generated by HPC modelling software. Modelling and curation are currently disparate. Part of the project work will include leveraging work on dcTARDIS to make the BDP seamlessly interoperable with HPC modelling tools (ie. to run on RMIT clusters, VPAC and NCI) such as VASP, CRYSTAL, SIESTA, GULP and DFTB. These are widely used across computational physics and chemistry.

As some of these tools are commercial-off-the-shelf (COTS), the BDP will be extensible by the community based on generic interoperation protocols.

Visualisations and Data-Driven Publications:

Data modelling visualisation applications will be integrated and enhanced within the MyTARDIS interface, providing a seamless context for hosted data. Researchers will be able to visualise, annotate and share representations of protein structures, diffraction images and other 2D and 3D representations within their browser.

The web-based application Jolecule (http://jolecule.appspot.com/) will be integrated into MyTARDIS and its annotation and collaboration capabilities enhanced. Jolecule is a 3D protein viewer designed by a Monash software developer (Dr Bosco Ho), which allows the visualisation of biological macromolecules in modern web-browsers. Jolecule is designed to be integrated with a central server, which allows users to annotate a given protein, providing a "wiki" type interface for sharing complex structural information.

These rich representations of data will be used as a strategy to generate reports and the basis of scientific papers, aiding scientists in describing their scientific workflow. Data-driven publications will be richly interactive and collaborative.

One particular area that this approach will impact on is particularly noteworthy - protein crystallography. There is currently intense debate in this field on the urgent need for publication of raw data. The Buckle Lab, in collaboration with the Monash eResearch Centre have led the way in the past 3 years via the TARDIS and MyTARDIS project, which is now being adopted by the International Union of Crystallography and several synchrotrons overseas (eg the Advanced Photon Source, IL, USA). The proposed work will deliver several key solutions that will be adopted worldwide by journals and institutions: (1) large data-centric portions of a publication will be automatically generated by MyTARDIS in a standardised format (eg data processing statistics); (2) data (both raw and derived) will be available in a secure format to journal reviewers such that sensitive information can be validated but not exploited (this is currently impossible and a burning issue in the field); (3) raw data and workflows will be published alongside the scientific report - this is not possible at present. The Buckle laboratory is collaborating with the Public Library of Science (PLoS) and Prof Philip Bourne (UCSD, US) in the area of data-centric publishing. This proposal will integrate closely within this collaboration; (4) Raw data will be available for the development of novel image processing technologies – access to such data on the proposed scale is currently impossible and is a significant driver of algorithm development.

The BDP's data driven publications will leverage the several ANDS services such as Register My Data and Identify My Data. Should the proposed ANDS Cite My Data service be put into production, then the assignment of Digital Object Identifiers would be made possible within the platform, allowing a richer description of data for citation.

RESEARCH COMMUNITY NEEDS & BENEFITS			
9	**Target Research Community**	*Identify the research communities and the expected number of users.*	0.5

The following section lists key researchers and their research groups who will lead the development of the Bioscience Data Platform.

Structural Biology: A/Prof. Ashley Buckle's laboratory

- 2 Postdocs, 4 PhD Students

Computational Biophysics: Prof. Dougal McCulloch, Prof. Salvy Russo, Prof. Ian Snook

- 5 postdocs, 10 PhD students

In addition to these principal research groups, this proposal is intended to compliment the Characterisation Virtual Lab proposal. Should both proposals be successful, this tool will be integrated into the Structural Biology Workbench of the CVL Fabric, providing a data repository to the data being generated and used within. The Structural Biology Workbench's lead research group is that of Prof. James Whisstock at Monash University.

10	**Needs and Impact**	*Describe the needs of the research community. Describe the impact on current research practices and any opportunities the new infrastructure will provide to those communities.*	2

		List the benefits to be derived from the delivered infrastructure, describe them in quantifiable terms where possible. *Outline how the impacts and benefits will be tracked, managed and measured.*	

Structural Biology:

Attacking important problems in structural biology requires the integration of several key techniques, notably Small-Angle X-ray Scattering (SAXS), and Cryo-Electron Microscopy (Cryo-EM). For example, crystallography is typically applied to molecules of molecular mass << 1 million Daltons (MDa), whereas SAXS and Cryo-EM can be used for molecules orders of magnitude larger. However, such approaches typically produce low-resolution data (5-15 Å, compared with 2-4 Å from X-ray studies) and this presents enormous challenges for building atomic models. To infer structure from such instruments requires the use of complex and sophisticated models that bring together different kinds of experimental data not only from structural, biophysical and biochemical methods but also from systems biology to define complex networks of interactions.

This proposal focuses on developing the computational technologies required for this integrated approach. It builds upon existing expertise in computational and structural biology at Monash, particularly in protein crystallography, structural bioinformatics and molecular modelling, in addition to existing collaborations with scientists at the nearby Australian Synchrotron. Further, Monash has electron microscopy facilities on campus, and the creation of the Clive and Vera Ramaciotti Centre for Structural Cryo-Electron Microscopy facility at Monash in 2012.

This proposal will serve as a foundation on which additional projects will build, leading to ARC and NHMRC project proposals and the potential to lead ultimately to NHMRC Program proposals and ARC Centre proposals.

Computational Biophysics:

RMIT houses a large Microscopy and Microanalysis Facility (MMF) opened in 2010 by Senator Kim Carr. The Bioscience Data Platform will allow raw data publication directly from the MMF including computational models derived from or validated using these data and employing HPC facilities.

The strength of the RMIT group is in computational biophysics upwards from the quantum level, combined with nanostructure, biophysics and biochemistry characterisation to feed into accurate modelling and predication of functional properties and processes. The data generated is massive and the computational resources required put RMIT usage of national and state HPC resources in the top 3 by compute hours routinely. Research practices in this group are increasingly data-centric for this reason and well beyond traditional curation of data collections.

The core group involved with this project includes Profs Russo, McCulloch and Snook, who work with 5 postdocs and over 10 PhD students on self-assembling nanosurfaces and the functional characterisation and synthesis of nanostructures. Over the past 10 years Russo and

McCulloch alone have held as sole CI or co-CI: 4 DP grants (DP0986713, DP0771646, DP0666866, DP0452796); 2 LP grants (LP0776931, LP0668091); a few industry contracts; 11 LE grants dominantly involving characterisation equipment (LE0989615, LE0882821, LE0882246, LE0775562, LE0668091, LE0668381, LE0560679, LE0454166, LE0453732, LE0346968, LE0231228). These grants total over $10.5mill in public funding. The proposed project will leverage this funding for the broader characterisation community of biophysicists and biochemists across Australia.

Given the moderate size of RMIT internal clusters, Profs Russo and McCulloch rely on a range of external facilities incl. VPAC, VLSCI, NCI and in the future the NeCTAR Research Cloud. A major problem is the heterogeneity of the HPC this community typically depends on. A more seamless interoperation with HPC centred on and controlled from the data itself stands to accelerate research significantly and reduce compute times from weeks to days.

Also importantly, driving the necessary HPC directly from the data models will decrease the threshold for young researchers entering this field, as increasingly more idiosyncratic HPC control details can be hidden from them and they can focus on the underlying biophysics and biochemistry in a model-driven approach.

Utilisation and uptake of the BDP and its increase over time will be measured via these methods:

- Web traffic analytic software installed for the number of page hits and their originating source location.
- An internal mechanism in MyTARDIS to track and record user signups, logins, downloads and HPC job scheduling.
- The number and location of scientific instruments / HPC resources being linked to the BDP and the quantity of data stored for access.
- The use of ANDS services for registering data in the BDP for Research Data Australia.

11	**Broader adoption**	*State which additional communities, resource providers or organisations would also be expected to benefit from the use of the new infrastructure and services should the project succeed.*	0.5

The Bioscience Data Platform is intended to be a freely available service to all Australian researchers. As the platform's core will be MyTARDIS, the platform will already have the capability to receive all data collected at instruments that are MyTARDIS-enabled. This currently includes 4 instruments at the Australian Synchrotron (Macromolecular Crystallography 1 and 2, Infrared, Small/Wide Angle X-Ray Scattering) and at ANSTO, with 5 beamlines at the Bragg Institute linked to MeCAT (based on MyTARDIS). Pending the success of the Australian Synchrotron's NeCTAR Research Tool proposal, this would enable the nationally deployed Bioscience Data Platform to receive data from all instruments at the Australian Synchrotron and ANSTO. This would be available to all Australian researchers as a cloud service, meaning zero deployment or configuration requirements.

Similarly, existing and upcoming MyTARDIS installations for microscopy/microanalysis (RMIT, University of Queensland) and physics (University of SA, RMIT) will have the capability to push their experiments to the Bioscience Data Platform for further analysis, collaboration and publication.

12	**Value adding**	*Identify the components of the project that are adding value to existing research infrastructure investment. For example; building new services on top of and using existing research infrastructure.* *Describe the alignment with national research infrastructure and eResearch infrastructure priorities. See http://www.innovation.gov.au/Science/ResearchInfrastructure/Pages/default.aspx.* *Note how this project will engage with, and leverage off, the other national infrastructure programs such as ANDS, AAF, NCI, and RDSI, as well as any other NCRIS Capability Platforms.* *Identify components that could be used by other research communities and organisations to resolve their problem. For example; a part of the workflow that could be used by other domains and research communities.*	1

The Bioscience Data Platform builds on the MyTARDIS data management system. MyTARDIS has been enriched by several ANDS-funded projects such as:

- Metadata Store (EIF019): Squirrel (Monash University)
- Data Capture projects:
    - Bioscience Data Platform (Monash University)
    - MicroMyTARDIS (RMIT)
    - dcTARDIS (RMIT)
    - MeCAT (both Australian Synchrotron and ANSTO)

The BDP will be leveraging these technologies:

- Implementing the Australian Access Federation as the BDP's main method of authorisation
- Integration with aspects of national imaging and visualisation facility MASSIVE:
    - Structural Biology Tools (including Visualisation and HPC)
    - GPU resources for faster, convenient visualisation of data in the cloud
- Engagement with VPAC and NCI-NF for the running of simulation and molecular dynamics software via the BDP, such as:
    - VASP
        - A Density Functional Theory program for simulation of material properties using plane waves.
        - Used by over 45 researchers from RMIT University, University of Sydney, University of Melbourne, University of Queensland, CSIRO, Monash University, ANU, Newcastle University, Curtin University

  - GULP
    - A Density Functional Theory program for simulation of material properties using Gaussian basis sets.
    - Used by over 10 researchers from RMIT University, University of Sydney, CSIRO, ANU, Newcastle University, Curtin University
  - SIESTA
    - A Linear Scaling program for simulation of material properties.
    - Used by over 25 researchers from RMIT University, University of Sydney, University of Melbourne, University of Queensland, CSIRO, ANU, Curtin University, University of Tasmania
  - CRYSTAL
    - Empirical potential program for simulation/modelling of material properties.
    - Used by over 10 researchers from RMIT University, ANU, University of Melbourne
  - PHASER
    - Macromolecular crystal structure phasing program with maximum likelihood methods.
    - Used by over 50 researchers from Monash University, University of Melbourne, University of Western Australia, University of Queensland, University of Sydney
- Implementing ANDS Services
  - Register My Data
  - Identify My Data (handle.net)
  - Cite My Data (DOI minting, pending official commissioning by ANDS)
- Integrating and enhancing the Jolecule 3D protein viewer
- The advanced capabilities hosted at Monash under both the Research Data Storage and Infrastructure (RDSI) and the NeCTAR Research Cloud
- Working with the technologies described in the Characterisation Virtual Lab's Structural Biology driver.

As described under 'Broader Adoption', the BDP will be available to all Australian Researchers. Further efforts can be made to link additional instruments to the BDP. The proposed modular 'app' structure for HPC integration and visualisation will allow support for further applications to be built into the platform, available to all.

PROJECT MANAGEMENT			
13	**Governance**	*State who is accountable for assessing project performance, what process will they apply.* *Describe the authority structure over resources in the proposed project.* *List all members of the Project Governance Body.* *Describe the organisation's Project Management methodology, scaled as appropriate for the proposed sub-project, and the maturity of its use within the organisation.*	1

		Describe the key processes and templates used internally by the organisation for governance and project management.	

Development of the BDP will be managed as a project by:

- Key Personnel (listed in item 7) who will be assessing ongoing project performance to ensure that developed infrastructure is in line with their research goals.
- A project manager that will coordinate the development across the two participating institutions, Monash University and RMIT to ensure a cohesive and compatible outcome.
- The BDP steering committee will be meeting at least quarterly.

The BDP steering committee has these confirmed members:

- Steve Androulakis (Project Manager, Monash University)
- A/Prof Ashley Buckle (Medicine, Monash University)
- Prof Heinz Schmidt (eResearch Director, RMIT)
- Prof Salvy Russo (Science, RMIT)
- Dr Stephen Crawley (AMMRF/UQ)
- Dr Ulrich Felzmann (Principal Computing Scientist, Australian Synchrotron)
- Paul Bonnington (Non-voting Chair)

Technical advice on the steering committee will be provided by Steve Androulakis and Ulrich Felzmann – both who have had a large involvement in the development/architecting of MyTARDIS and its connection to bioscience resources.

NeCTAR's director (or a delegate) is invited for membership to the BDP steering committee in an ex-officio capacity.

Governance of the BDP's core software, MyTARDIS is in line with Open Source practices. Ashley Buckle (Monash University) provides the central vision for the code base, while Steve Androulakis (Monash University) is leader of the open source community development.

Over eighteen months, more than fifteen software developers from four universities and five organisations have contributed code to the Google Code-hosted codebase. The software is freely and publicly available under a BSD license, and a Google Groups mailing list exists for community-support. Any developments of the BDP that adds to or changes the internals of MyTARDIS will be considered for merging into the central codebase.

The nominated project manager will have sufficient authority over the project resources to fulfil the role as described in section A-4.3.

14	**Project Scale**	*Identify the overall scale expected in the project, total effort, amount of funding required, amount of co-investment proposed and nominate any other participants that have indicated a willingness to participate through providing resources and what they are.*	0.5

The key participants are requesting $365,400 from the NeCTAR RT program for the development of the BDP.

The partners will contribute the following:

- 3 FTE of in-kind people ($392,309) for the duration of the project

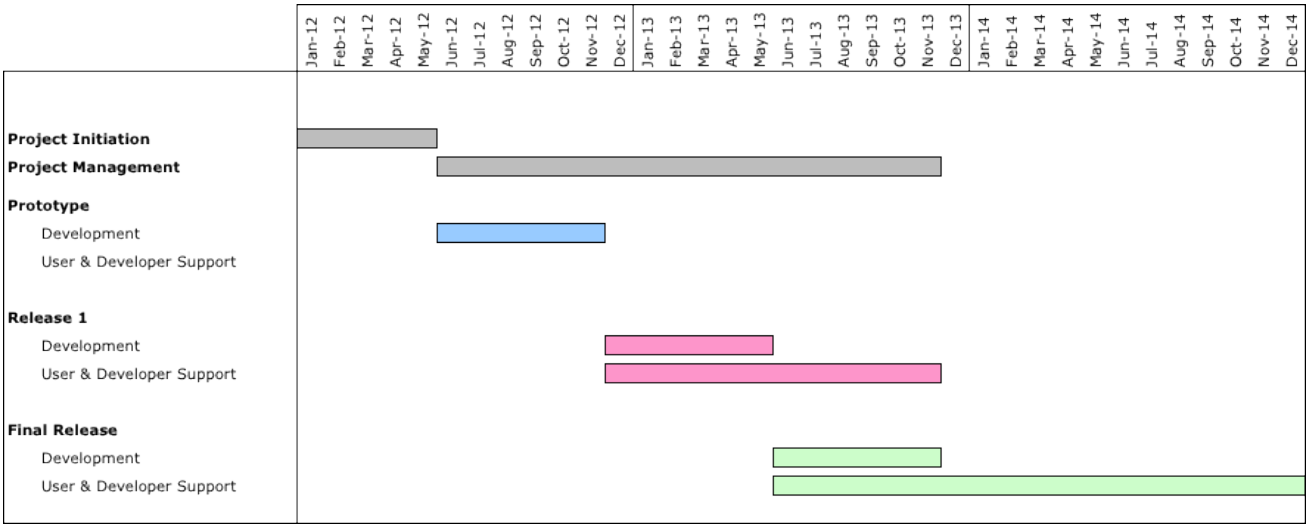
In total, the endeavour will manage 5.9 person-years of development, project management, support and operation between the start of 2012 and the end of 2014.

15	**Project Approach**	*Detail how the required services will be developed and delivered. Outline the different stages of activity.*	1

The product will be developed using an Agile Software Development methodology (Scrum), with the relevant researcher very much involved in the process and actively steering the product. Development and delivery will take the following approach:

- Collect initial requirements, conceptualise a solution, and produce a product backlog (for Scrum). Requirements will be further developed during development sprints.
- Design initial architecture. Architecture will be further developed during development sprints.
- Plan releases in a road map
- Plan iterations for Scrum
- Software development iterations (based on scrum and guided by the road map). Each iteration will be: incremental, include software testing, and further refine requirements & architecture.
- Software will be developed in an environment similar to production, in order to simplify and expedite deployment.
- At significant and regular stages (as guided by the roadmap), the developed software will be placed into production. This will deliver value to the researchers early and often and promote adoption. Product deployment will be conducted by eResearch application operational staff, not the developers, to enable the developers to focus on software development. They will also support the product.
- On completion of the iterations, formal training material will be completed and the product formally launched.
- During the entire process, the product will be promoted to the relevant research community and external collaboration and adoption promoted.
- Ongoing support for the product will continue with eResearch application operational staff.

The different stages of activity are outlined as follows:



16	**Key Deliverables and Acceptance Criteria**	*Define the key project deliverables.* *Deliverables include the services, infrastructure and functionality specified above for development by the sub-project as well as any required project management artefacts.* *Show the Acceptance Criteria against each deliverable. These will be further elaborated during project delivery when Commissioning Tests are prepared.* *Define the Acceptance Criteria specific to Commissioning.*	1

Project Area	Deliverables	Milestone	Acceptance Criteria	Interested Parties
BDP On NeCTAR Research Cloud	1a: Design and develop OpenStack connectors for MyTARDIS	3 months	Demonstration of the ability to upload, download and remove data from the platform.	eResearch community
	2a: Develop the ability for the BDP to receive feeds of data from MyTARDIS-enabled instruments and facilities	6 months	Demonstration of the BDP automatically receiving data and metadata from a remote instrument	The drivers
	3a: Linking to feeds of all available data from all current MyTARDIS installations	9 months	Demonstration	Community and the drivers
	4a: Australian Access Federation Integration	12 months	Demonstrate AAF authentication in the BDP with full group-management support and data sharing	Community and the drivers

	5a: Stable BDP automatically receiving data on NeCTAR Research Cloud	15 months	Demonstration and performance metrics	Community and the drivers
Integration of an enhanced 3D protein structure viewer, Jolecule into the Platform	2b: Jolecule's annotation functionality abstracted and compatible with MyTARDIS' internal metadata model.	6 months	Demonstration	The drivers
	2c: Development of an app within MyTARDIS to enable Jolecule to function within the BDP	6 months	Demonstration	Community and the drivers
	5b: Enhancement of Jolecule's annotation capabilities to support collaborative discussion and sharing of points of interest	15 months	Demonstration	Community and the drivers
A data-driven publication interface	2d: Links to domain-specific databases to disseminate information formatted for submission to publications	6 months	Demonstration	Community and the drivers
	3b: Development of a rich interface format for the display of metadata and visualisations stored within the platform	9 months	Demonstration and real-world data publication for citation in a scientific journal	Immediate research community: Centre for Structural Cryo-Electron Microscopy
	5c: Compatibility with the ANDS Register My Data service	15 months	Registration of an entry in ANDS' Australian Research Data Commons	Community and the drivers
	5d: Compatibility with the ANDS Identify My Data service	15 months	Demonstration of handle.net identifier minting	Community and the drivers
	5e: Compatibility with the ANDS Cite My Data service	15 months	Demonstration of DOI minting	Community and the drivers
An interface to High Performance Computing resources	2e: A common set of functionality to schedule jobs to High Performance Computing resources and retrieve their results	6 months	Demonstration of the ability to schedule jobs and retrieve their results within the BDP	eResearch community

	4b: Support for the setting of parameter sweeps and the rich visualisation of results for nominated applications	12 months	Demonstration of the scheduling of chosen applications on HPC resources and their rich visualisation	Community and the drivers
	6a: Capture of parameters for the rescheduling and publish of data workflows	18 months	Demonstration of the ability to re-run workflows, with changes	Community and the drivers

At each stage described above, a representative of the interested parties will be assessing deliverables and signing off on acceptance tests.

17	**Quality Control**	*Identify the personnel, processes and any special resources that will be required for Quality Control, Acceptance and Commissioning Testing activities on the proposed project.* *State who is responsible for the completion of each deliverable.* *State who is accountable, within the sub-contracted organisation, for the acceptance of each deliverable.* *In contrast to "what we will test" described earlier, this is a "how we test" description.*	1

To ensure a quality product, the software developers will employ continuous integration by:

- Maintaining a code repository
- Automating the build
- Making the build self-testing, using unit tests and test harnesses
- Everyone committing to the central repository every day
- Every commit (to the repository) being built
- Regular performance metrics performed on builds
- Testing in a clone of the production environment
- Automating deployment

Software quality will also be assured through peer reviews and system & integration testing by software developers who are not involved in developing code for the product.

Deployments of the product and initial release testing will be completed by eResearch application operational staff.

The project's project manager will be accountable for completion of each deliverable.

The researcher championing the product development will be accountable for the acceptance of each deliverable.

18	**Risk and Issue Management**	*Define the key risks to the successful delivery of the proposed project.* *Define any open issues that need resolving before the proposed project can start delivery.* *Detail any key questions that will affect the operation of the proposed project pending a decision.* *Define how the major risks and issues to the proposed project will be managed.*	1

A far-reaching project such as the Bioscience Data Platform has inherent risks associated with expertise, infrastructure and other areas.

Much of this platform relies on NeCTAR Research Cloud and high performance computing resources and environments. The Characterisation Virtual Lab accompanying this Research Tool is intended to be the host for much of this infrastructure. Without it, access to vast amounts of storage and computational power is hampered. Furthermore, early and consistent access to the NeCTAR infrastructure is seen to be critical to the success of this project. Similarly, adequate support from providers of NeCTAR Research Cloud infrastructure is seen as key.

The Bioscience Data Platform's core software infrastructure will be MyTARDIS, which is developed in Python and compatible with Python 2.x. The Australian Access Federation currently doesn't have a Python API, so at this point it isn't clear how much effort it will be to integrate this functionality. Possible solutions to this is the development of a Python API within this project, which will then be contributed to the community, or a separate, but complimentary module developed in an AAF-supported language deployed alongside as a gateway to the Bioscience Data Platform. Key personnel in this proposal have the skills necessary to achieve both of these ends.

Applications chosen to be integrated into the Bioscience Data Platform are likely to require specific expertise. Harnessing researchers' knowledge of specific software programs will require them to be available to us for amounts of time. However, this project aims to select widely popular software tools both for maximum impact and also vast support from user communities. Opinion can also be crowd-sourced for implementation details in an effort to get the widest range of opinions on operation of tools within the BDP.

One key piece of software to be integrated, Jolecule was developed by a programmer at Monash University. While the code is open source, one specific programmer developed it in its entirety. It's seen as a risk for one person to hold all expertise on a piece of software. Therefore early priority will be given to working with the programmer in an effort to gather knowledge on its internal operation. In a worst-case scenario, the open-source code base can be investigated by our software developers to gain needed insight.

LEVERAGING			
19	**Standardisation and Interoperability**	*Describe the global technology development or standardisation work that will be adopted, adapted or extended within the project and any risk reduction or additional value available by collaboration with similar activities occurring elsewhere in the world.* *Identify any local or emerging standards that will incorporated by the project.*	0.5

Although MyTARDIS has its origins in Monash University has become widely extended and deployed by an open-source community of developers.

Over eighteen months, more than fifteen software developers from four universities and five other organisations have contributed code to the Google Code-hosted codebase. The software is freely and publicly available under a BSD license, and a Google Groups mailing list exists for community-support.

MyTARDIS is built on widely used technologies such as Python, Django and JQuery. Python is a modern, popular programming language in both science and the web. MyTARDIS' python code is compliant with the PEP8 standard library.

Django is Python's most popular web framework. Django runs on the ubiquitous and free Apache web server and supports major databases such as PostgreSQL, Oracle, MySQL and SQLite. JQuery is one of the most popular libraries for javascript and it will continue to assist MyTARDIS and the proposed Bioscience Data Platform alike in providing streamlined rich interfaces.

MyTARDIS uses the METS XML standard as its ingestion and dissemination format.

FINANCIAL			
20	**Budget Breakdown**	*Provide a proposal breakdown of the project budget by the milestones, which are described as an attachment to the Proposal in the format described below.* *Include proposed staffing levels; where actual individuals have not been allocated to the sub-project, use a role name and Full Time Equivalents (FTE) to show the number and value of budgeted staff that will be working on the sub-project at that milestone. For example; Software Designer by 2 or two Software Designers working full time on the sub-project. Each individual or FTE role is to be included as a separate line item.* *Break down the budget into EIF (NeCTAR) funding or co-investment funding.*	1

SERVICES AND SUPPORT

21	**Service Levels**	*Specify the service level that will be offered for each service in relation to the levels discussed in the relevant Part B document.*	0.5

An accurate picture of service levels will be more achievable when more information is known on NeCTAR Research Cloud availability and service levels.

Nonetheless, the Bioscience Data Platform aims to achieve uptime of greater than 99%.

Support requests specific to the operation and use of the BDP will receive a response within 1 business day. Requests will be internally managed by a job tracking system (such as Jira) that allows priority to be assigned. This support includes new user activation, if manual activation is needed.

22	**Operations and User Support**	*Detail the proposed operator of each service, what support will be provided to users and by whom.*	1

The Bioscience Data Platform will be hosted and supported until at least 2014.

Departments at supporting institutions, the Monash eResearch Centre and RMIT's IT services group will provide hosting and technical support until the end of 2014.

0.5 FTE will be dedicated by Monash University under this project to the continued operation and upgrade of infrastructure from time of production until 2014 and beyond.

Both RMIT and Monash University will host aspects of the BDP in production. The BDP may take advantage of the hosting provided to the Characterisation Virtual Lab of which Monash is the host institution.

Any technical enquiries regarding the development of the platform as a whole can be directed to the MyTARDIS Open Source community. A google group, MyTARDIS-devel exists for the answering of queries specific to the development of the core software.

Given the amount of scientific expertise required to use of many of the BDP's services, there exists a need for scientific user support. It's also acknowledged that those wishing to extend the suite of tools available within the BDP will need specific technical information related to implementation. Both the Monash University eResearch Centre and RMIT University will provide this level of support – diverting requests to expertise in relevant faculties where necessary.

23	**Sustainability**	*Describe how the project infrastructure will be made sustainable following the completion of the project and becoming operational, when EIF funds cannot be used. Include information on the proposed business/financial model and the timeframes to which they apply.*	0.5

The sustainability of the Bioscience Data Platform is seen as being achieved through the continual use and enhancement of its open source code base, and a critical mass of users interfacing the platform. The BDP's core, MyTARDIS has a healthy open source community based on its extensibility and ease of application to many areas of research.

Components of the BDP such as the HPC connector interface and rich data publication components will be written in an abstract, modular fashion. It's foreseen that these components will be taken and remodelled for new connections and integration with new applications and resources. This will see the suite of tools available to the BDP updated and replaced as the science drivers change and evolve over time.

24	**IP, Licensing and Access**	*State if any software licenses will be used for software developed by the proposed project, or other software used for the services to be delivered by the proposed project.* *State if software developed by the project will be made available under an open-source license.* *State if other Intellectual Property (IP) or licensing restrictions are relevant to the services that will be delivered.* *State if there are any restrictions on access to the services that will be delivered.*	0.5

As far as practicable an open source approach will be taken to software and services developed in this project. In some cases the ability to do this may be limited or may require consents from owners of software required to implement aspects of platforms or tools that interact with those to be developed here. Examples include where data acquisition, transfer and/or initial processing requires the use of third party software under license.

Any software that requires commercial licensing such as proprietary scientific software tools will be involved in a periphery capacity only. These tools may be interfaced with by the BDP, but will not form part of its software or development efforts.

The software developed in this project will be made available under an open-source license – again as much as is possible dependent on the inputs.

25	**Communications and Engagement**	*Describe the means by which customer satisfaction with the proposed project's planning, requirements gathering, scoping decisions, progress, quality and outputs will be measured.*	0.5

The BDP has been developed in close engagement with the key research personnel who stand to benefit from this infrastructure. These close links will continue through the project's duration and beyond.

It's planned that the BDP will be governed and managed in close collaboration with the Characterisation Virtual Lab's Governance and Management bodies. This will help to ensure that the project's focus, scope and success are tied to key scientific drivers.

Regular testing of functionality will be performed with research groups, and their feedback continually incorporated into the development process.

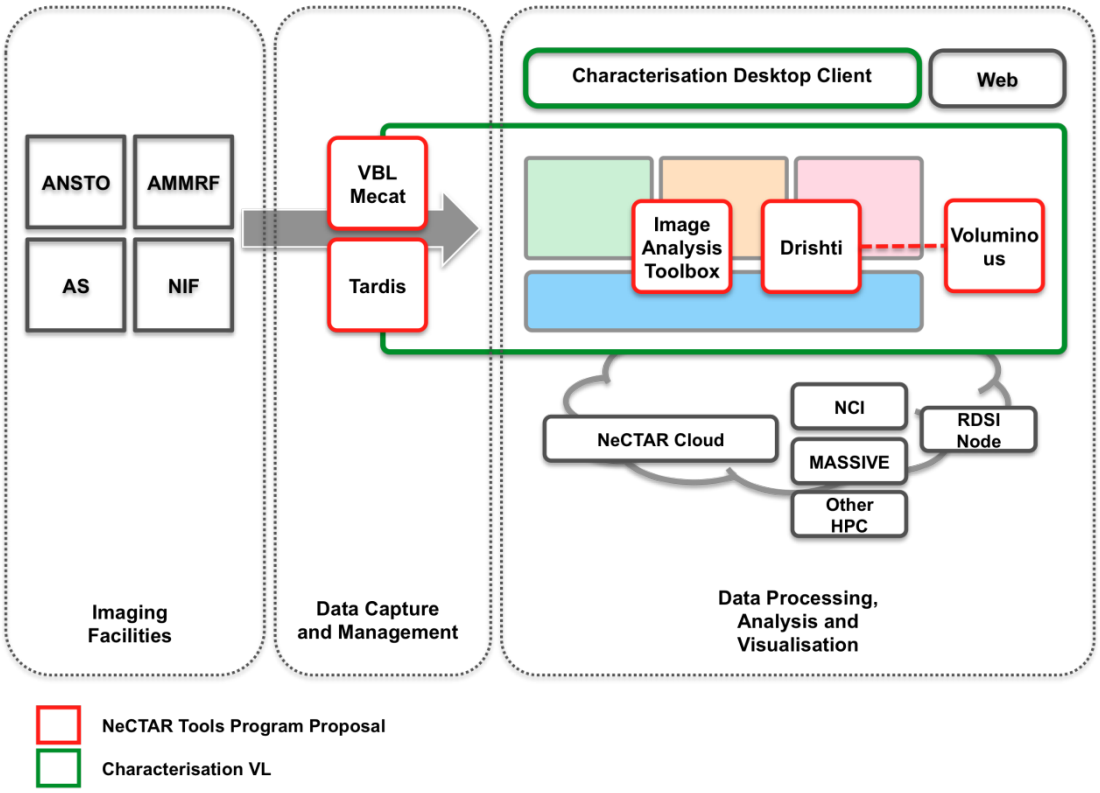
The research communities involved in the development of the BDP will take ownership over the solution and present their achievements to their collaborators via scientific publications and conference papers.

26	**Constraints and Dependencies**	*Define and explicitly quantify any schedule, expenditure, resource, scalability, performance, and quality constraints or limitations on the project and its deliverables.* *State the dependencies with external parties, including other NeCTAR projects, which have been identified in planning the proposed project.*	0.5

The BDP project depends on:

- The NeCTAR Research Cloud infrastructure and the NeCTAR Characterisation Virtual Lab project.

The BDP can function and provide benefit without the Characterisation Virtual Lab, however its development will be hampered by the lack of core software infrastructure the CVL would provide. How the BDP (shown as TARDIS) would function within the Characterisation Virtual Lab infrastructure is shown below.



The Monash University-led Characterisation Virtual Lab's Structural Biology Driver will be providing virtual environments containing tools and linkages to high performance resources. It will be seen that many of the tools hosted in these environments will be of benefit to the bioscience community within the BDP Research Tool. To leverage this, the BDP will be able to directly interface with the environments and resources hosted by the Characterisation Virtual Lab. Similarly, the inverse is true: Users of the Characterisation Virtual Lab will have a simple pathway for cataloguing and sharing their data via the BDP.

The BDP will be hosted as a nationally accessible service on Monash University's NeCTAR research cloud infrastructure.

External Party	**Capability Required**	**Date First Required**	**Milestones or Deliverables dependent on that capability**
Characterisation Virtual Lab	Ability to launch CVL Environment via browser	10 March 2013	C (3b)
NeCTAR Research Cloud	Able to host BDP on NecTAR Research Cloud as first release to community	10 June 2013	D
RDSI	*None Known*	-	-
AAF	*Capability (Web Services for*	-	-

	gateway) Currently Exists		
MyTardis-enabled Instruments	A pathway added to existing MyTardis-enabled instruments to push their data to the BDP (an administrator action per instrument)	10 March 2013	C (3a)
HPC Facilities	Access to compute time on Monash HPC and RMIT-accessible NCI HPC resources for testing and development	10 December 2012	B (2e)

For National Servers Program and Research Cloud proposals Only – *Not Applicable*			
27	**Datacentre**	*Provide information about the datacentre hosting the infrastructure and the existing capacity for the projected equipment. Define what emergency and redundant power is available State if it is cooling redundant, and with sufficient capacity. Define what fire systems are used. State your stated service levels for your datacentre.*	0.5
28	**Network**	*Provide information about the network facilities available to the infrastructure. Define the typical network capacity available to servers in your datacentre. State any known bottlenecks (eg in the network core). State your connection to AARNET. Define any other wide area network links that you have that could be useful to this project. State if you configure critical network equipment in a redundant manner. State if you have stated service levels for your network.*	0.5
29	**Operations**	*Provide information about your ICT operations practices that would apply to this project. State if you have maintenance windows, or procedures for handling infrastructure downtimes. State what monitoring facilities are used. State your typical hours of operation.*	1
30	**Governance and Participation**	*Indicate your commitment to participating in the RC/NSP Governance framework as outlined in the relevant Part B, working cooperatively with the Lead Node and the NeCTAR governance groups to achieve a nationally coherent infrastructure*	0.5
31	**Research Application Migration**	*Indicate if you are applying for funding to support the migration of Research Applications to the Research Cloud, and the level of funding requested (maximum $500K). Broadly describe the nature of the application migration activities that are proposed to be undertaken by the Node, including where special support will be provided to particular research domains, or particular classes of*	1.0

		research applications.	

Section 6 Selection Criteria

The following table outlines the criteria that will be used to assess proposals, based on the responses to Section 5 above. They are provided here only for the information of respondents and to ensure that responses consider the key elements being sought.

6.1 Criteria for Virtual Laboratory and eResearch Tools Proposals

Category	Weight (%)	Criteria
Research Community	**20%**	
Research community to benefit is well-defined (by location, institutions, size), nationally significant and the proposal is well supported by the research community		
The needs of the research community to be addressed by the proposal are well defined and significant		
Researcher participation in the proposal is well defined, significant and national or international in scale, supporting outreach, uptake, testing and evaluation of the project infrastructure		
The proposal is aligned to national research priorities		
The proposal is aligned with, and contributes to implementation of, national research infrastructure and eResearch priorities		
Research Impact	**30%**	
The benefits to be delivered to the identified research community are well described, achievable, significant and measurable		
The process for tracking and measuring the benefits is defined and achievable		
Virtual Laboratories:		
The proposal integrates significant infrastructure and research capabilities on a national scale to deliver a transformative impact for the research community identified in the proposal: • Improving access to instruments, data, compute and other research infrastructure • Enables new research practices through research workflows • Address emerging research challenges • Support cross-institutional and cross-disciplinary research workflows through the provision of integrated collaborative ICT infrastructure • Connect significant infrastructure capabilities to support discipline and problem oriented research workflows: e.g. remote laboratory access, computation, research data repositories, workflow tools and sensor networks • Provide an exemplar to research communities of the benefits of integrating significant research support capabilities into a rich online collaborative environment.		
eResearch Tools:		
The proposal delivers software infrastructure which improves existing tools to: • Enhance support for research collaboration • Improve remote access to underlying research facilities and infrastructure • Enhance support for research workflows, including cross-institutional and cross-disciplinary workflows, and/or • Enable the connection of research data sets and repositories with research tools and workflows.		

Potential benefits from re-use across research disciplines are well described, realisable and significant		
Implementation	**30%**	
An appropriate Governance structure has been defined for the Project • A Governance body appropriate to the project has been defined • Authority structure over resources in the project has been described • Key personnel (if required) and their roles have been clearly defined		
The infrastructure to be created by the proposal is well-described, achievable and will deliver the research impacts described. • The scope is realistic and outcomes are achievable • Project Management is well-described and appropriate to the proposal scale • Key risks have been identified and are manageable • Issues that require solving have been identified • Dependencies with third parties have been listed		
Capability and track record in development and operation of eResearch infrastructure • The contribution and track record of each organisation in the proposal to development and operation of the infrastructure is well-described and appropriate to the successful delivery of the proposal infrastructure		
The proposal leverages or builds upon existing research infrastructure where appropriate: • the NeCTAR Research Cloud and NSP for infrastructure hosting and computation • existing research and eResearch infrastructure (eg. RDSI, ANDS, NCI, Pawsey, Super Science, NCRIS Capabilities, State and Institutional infrastructure) It is expected that proposals will utilise the AAF for common authentication services. Where it is not possible for the proposed infrastructure to utilise the AAF for authentication purposes, appropriate and reasonable justification has been provided.		
Financial and Co-investment	**20%**	
The project budget is well-described, matched to appropriate milestones and appropriate to the needs of the project		
The identified co-investment achieves the target level, is appropriate to the needs of the project, and adequately covers the operational requirements of the proposal		
The proposal identifies an appropriate model for delivering future sustainability of the infrastructure		
Proposed expenditure of EIF funds is adequately described and conforms to the EIF funding guidelines (Mandatory)		
The proposal conforms to the principles on Access and Pricing as described in Part B of the NeCTAR RFP (Mandatory)		

6.2 Criteria for Research Cloud and National Server Program Proposals

Category	Weight (%)	Criteria
Research Community	**10%**	
Research community to benefit is well-defined (by location, institutions, size), nationally significant and the proposal is well supported by the research community		
The needs of the research community to be addressed by the proposal are well defined and significant		
Researcher participation in the proposal is well defined, significant and national or international in scale		
The proposal is aligned to national research priorities		
The proposal is aligned with, and contributes to implementation of, national research infrastructure and eResearch priorities		
Research Impact	**20%**	
The benefits to be delivered to the identified research community are well described, achievable, significant and measurable • Process for tracking and measuring the benefits is defined and achievable		
Potential benefits from re-use across research disciplines are well described, realisable and significant		
Implementation	**40%**	
An appropriate Governance structure has been defined for the Project • Proposers commit to operate under the Governance arrangements for the RC and NSP as described in the NeCTAR Final Project Plan (Section 4.3) • Research Cloud node proposers commit to operate in accordance with the principles described in Section 4.3.1.3 of the NeCTAR Final Project Plan • Authority structure over resources in the project has been described • Key personnel (if required) and their roles have been clearly defined		
The infrastructure to be created by the proposal is well-described, achievable and will deliver infrastructure in a timely manner. • Project Management is well-described and appropriate to the proposal scale • Key risks have been identified and are manageable • Issues that require solving have been identified • Dependencies with third parties have been listed		

Capability and track record in development and operation of signifcant ICT infrastructure for research users • The contribution and track record of each organisation in the proposal to development and operation of the infrastructure is well-described and appropriate to the successful delivery of the proposal infrastructure
The proposal leverages or builds upon existing research infrastructure where appropriate: • Co-location with proposed nodes of the RDSI Project • Co-location of proposed Research Cloud and NSP nodes to achieve cost reductions • Co-location with other significant eResearch infrastructure (eg. HPC, Repositories Instruments)

Financial and Co-investment	**30%**	
The project budget is well-described, matched to appropriate milestones and appropriate to the needs of the project		
The identified co-investment achieves the target level, is appropriate to the needs of the project, and adequately covers the operational requirements of the proposal		
The proposal identifies an appropriate model for delivering future sustainability of the infrastructure		
Proposed expenditure of EIF funds is adequately described and conforms to the EIF funding guidelines (Mandatory)		
The proposal conforms to the principles on Access and Pricing as described in Part B of the NeCTAR RFP (Mandatory)		

Section 7 Milestone and Funding Milestone Template

7.1 Funding Estimate

Please add the details of any anticipated participating organisations in the below table along with their anticipated Funding Allocation as a percentage of the Proposer's Total Funding estimate.

Organisation / Group Name	Anticipated Distribution of EIF Funds (%)
Monash University	52%
RMIT University	48%

7.2 Milestone Template

Complete the table overleaf with proposed milestones and the associated budgets and proposed funding amounts to be drawn down from NeCTAR. The submitted table must form an attachment to the Proposal and will be used to prepare the contract Schedules. Deliverables are to be described in as much detail as necessary to show that careful thought

has been spent on planning. Example milestones shown may apply to a particular type of project, but are expected to be adapted to suit the needs of the project and NeCTAR Program.

Note – Items in "Deliverables/Completed Activity" are mandatory.

	Funding Milestone Yes / blank	Milestone Title	Deliverables/Completed Activity	Target Milestone Date	NeCTAR (EIF) funds ($thousands)				Co-investment (budgeted contribution value) ('000)
					Requested ('000)	Planned Expenditure breakdown			
						Labour ('000)	Equipment ('000)	Other ('000)	
1	Yes	Sub-contract signed		1 April 2012	183.141	183.141			98.077
2		Project Initiation complete	*Communications plan prepared and sent to NeCTAR (Signed contract + two months).*	30 April 2012					
3									
4		Agreements signed with the participants		10 June 2012					
5	Yes	A: Early Funding Milestone	1a: Design and develop OpenStack connectors for MyTARDIS	10 Sep 2012	48.838	48.838			78.462
6		B: Prototype developed for demonstration to drivers	2a: Develop the ability for the BDP to receive feeds of data from MyTARDIS-enabled instruments and facilities 2b: Jolecule's annotation functionality abstracted and compatible with MyTARDIS' internal metadata model. 2c: Development of an app within MyTARDIS to enable Jolecule to function within the BDP 2d: Links to domain-specific databases to disseminate information formatted for submission to publications 2e: A common set of functionality to schedule jobs to High Performance Computing resources and retrieve their results	10 Dec 2012					

7	Yes	C: Linking existing data connections to BDP and rich display of data	3a: Linking to feeds of all available data from all current MyTARDIS installations 3b: Development of a rich interface format for the display of metadata and visualisations stored within the platform	10 March 2013	48.838	48.838			78.462
8		D: First release to community	4a: Australian Access Federation Integration 4b: Support for the setting of parameter sweeps and the rich visualisation of results for nominated applications	10 June 2013					
9	Yes	E: Automated links from data sources, annotation, and connections to ANDS services	5a: Stable BDP automatically receiving data on NeCTAR Research Cloud 5b: Enhancement of Jolecule's annotation capabilities to support collaborative discussion and sharing of points of interest 5c: Compatibility with the ANDS Register My Data service 5d: Compatibility with the ANDS Identify My Data service 5e: Compatibility with the ANDS Cite My Data service	10 Sep 2013	48.838	48.838			78.462
10		F: Final Release	6a: Capture of parameters for the rescheduling and publish of data workflows	10 Dec 2013					
11	Yes	G: Final Admin Closure	*Post-implementation Review (PIR) conducted and sent to NeCTAR.* *Practical Completion Certificate accepted by NeCTAR.*	22 Dec 2013	36.628	36.628			58.846
12		Operations to June 2014	*Service Levels met and reported to NeCTAR as defined.*						

Milestone No.	Name Of Service / Deliverable	Date of Deployment for Pilot Use	Date of Deployment as Production Service
A	Early Funding Milestone	10 Sep 2012	-
B	Prototype developed for demonstration to drivers	10 Dec 2012	-
C	Linking existing data connections to BDP and rich display of data	10 March 2013	-
D	First release to community	-	10 June 2013
E	Automated links from data sources, annotation, and connections to ANDS services	-	10 Sep 2013
F	Final Release	-	10 Dec 2013