**Project:** National eResearch Collaboration Tools and Resources

**Project #:** 2179

**Contents:** Part A Brief

Part B Program Documentation

B1 Virtual Laboratories

B2 eResearch Tools

B3 Research Cloud

B4 National Servers Program

Part C Proposed Sub-Contractor Agreement

Part D **Proposal Submission *(this document)***

*Attachment 1 Research Cloud Specifications*

*Attachment 2 National Servers Program Specifications*

**Issue date**: 20th September 2011

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

Copyright of this document remains the property of The University of Melbourne

|  |  |
| --- | --- |
|  | 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](#_Toc303782264)

[1.1 RFP Contact Details 1](#_Toc303782265)

[1.2 RFP Timeline 1](#_Toc303782266)

[1.3 RFP Checklist 1](#_Toc303782267)

[1.4 Submission Instructions 1](#_Toc303782268)

[1.5 Late Submission 2](#_Toc303782269)

[Section 2 Contact Details of the Proposer 3](#_Toc303782270)

[2.1 Proposer Contacts 3](#_Toc303782271)

[Section 3 Compliance Statement 4](#_Toc303782272)

[3.1 Proposed Sub-Contract Compliance 4](#_Toc303782273)

[3.2 RFP Compliance 4](#_Toc303782274)

[3.3 Conflict of Interest 5](#_Toc303782275)

[3.4 Statement of Departures 5](#_Toc303782276)

[3.5 Conflict of Interest 5](#_Toc303782277)

[Section 4 Fields of Research 6](#_Toc303782278)

[Section 5 Response Template 6](#_Toc303782279)

[Section 6 Selection Criteria 32](#_Toc303782280)

[6.1 Criteria for VL and eRT 32](#_Toc303782281)

[6.2 Criteria for RC and NSP 34](#_Toc303782282)

[Section 7 Milestone and Funding Milestone Template 35](#_Toc303782283)

[7.1 Funding Estimate 35](#_Toc303782284)

[7.2 Milestone Template 36](#_Toc303782285)

# RFP

## RFP Contact Details

|  |  |
| --- | --- |
| **RFP Proposals ONLY** | [proposals-rfp-nectar@unimelb.edu.au](mailto:proposals-rfp-nectar@unimelb.edu.au) |
| **RFP Questions ONLY** | [questions-rfp-nectar@unimelb.edu.au](mailto: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 |

## 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 |

## RFP Checklist

|  |  |
| --- | --- |
| 1. Have you registered online at http://www.nectar.org.au? |  |
| 1. Have you read and understood Part A? |  |
| 1. Have you read and understood the relevant project Part B documentation? |  |
| 1. Have you read and understood Part C? |  |
| 1. Have you completed all sections of Part D? |  |
| * 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 |  |
| 1. Have you asked any questions you needed to, and received sufficient answers? |  |
| 1. Have you returned the pack, Part D, to [proposals-rfp-nectar@unimelb.edu.au](mailto:proposals-rfp-nectar@unimelb.edu.au)? |  |

## 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.

## 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:

1. resulted from the mishandling of the Proposal by NeCTAR; or
2. 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.

# Contact Details of the Proposer

## 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.

### Proposer

|  |  |
| --- | --- |
| Organisation Name | Centre for Genetic Epidemiology & Biostatistics, University of Western Australia |
| Contact Name | Paul White |
| Position | Manager Informatics Systems Development |
| Business Address | 35 Stirling Highway, Nedlands, WA, 6009 |
| Postal Address | 35 Stirling Highway, Nedlands, WA, 6009 |
| Telephone | 08 6488 6733 |
| Facsimile | 08 6488 6750 |
| Mobile Phone | 0412 352 883 |
| E-mail | paul.white@uwa.edu.au |

### Participating Organisations

|  |  |  |
| --- | --- | --- |
| Organisation / Group Name | Location | Role |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Compliance Statement

## 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

## 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

## 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

## Statement of Departures

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Clause or Reference** | **Nature of Compliance** | **Proposed wording of amendment** |
| **Proposed Sub-Contract** |  |  |  |
| **RFP Terms and Conditions** |  |  |  |

## Conflict of Interest

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

# 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) |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  | 0% |

# 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.

**SUMMARY**

# Program and Proposal Title (0.1 pages)

* *Which program is this proposal addressing? (RT, VL, RC, NSP).*
* *Provide a short title for the proposal to use as a reference in communications.*

Cloud-based Bioinformatics Tools

This proposal is addressing the eResearch Tools program.

# Executive Summary (1 page)

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

Data is the primary asset of biomedical researchers, and the engine for both discovery and translational medical research. As the volume and sensitivity of research data increases, due to technologies such as next generation sequencing of human specimens linked to clinical datasets, so too does the requirement for access to application software for integrating and interrogating the different types of research data. Researchers often need to execute complicated queries and conduct analysis across multiple data types, such as phenotypic, genotypic, pedigree and biospecimen data.

The Ark project was established within the Centre for Genetic Epidemiology & Biostatistics at the University of Western Australia in late 2009 to develop an open-source platform to support Australian and International biomedical researchers. The objective of The Ark project is to provide a suite of secure, integrated web-based applications that incorporate the majority of the functionality required to conduct a complex study or clinical trial.

The project has now delivered a Java-based solution that delivers the following capabilities:

1. Create and a configure a study;
2. Define users and manage their roles and access permissions
3. Define and manage a research participant pool;
4. Define and manage the collection of phenotypic research data;
5. Dynamically generate Electronic Data Capture (EDC) forms for collecting most types of textual research data;
6. Manage physical biospecimens and the associated data (LIMS); and
7. Generate reports.

This functionality has now been released as Version 1.0 and is being used in production. The project team is currently focussed on supporting Version 1.0 and gathering the requirements for the functionality described in this Proposal.

The software has been designed and developed in accordance with the following guiding principles:

1. The software should be web-based;
2. One instance of the software should securely support multiple research studies and multiple users;
3. The core application software should not be developed to suit any one study/project but rather should be configurable to suit a broad range of studies;
4. The software should be as user-configurable as feasible through a browser-based graphical user interface;
5. The software must be highly secure; and
6. The software modules should be well integrated.

This proposal has been developed with goal of implementing the additional functionality required by our collaborators and the majority of medical researchers currently undertaking data collection and analysis projects. In addition to enriching existing capabilities this functionality includes:

* Integration with the AAF authentication services;
* An integrated invoicing and billing module;
* A data extraction for analysis module;
* A pedigree management and visualisation module. Note that we have developed a prototype for a 3 dimensional pedigree visualisation tool that may be used as the basis for this module;
* Registry Management functionality for managing participant registries, such as The Australian Twin Registry; and
* A genotypic data management module. The preferred approach to providing this functionality is to integrate The Ark tools with the capabilities that would be provided through the Genomics Virtual Laboratory, Galaxy/GDR Integration and Service Centre Data Handover projects through a set of web services, should they be funded. Note that we have developed a genotypic data management prototype in conjunction with researchers and developers from the OBiBa project at McGill University in Montreal that could form a basis for this capability should the other projects not be funded. In the event that these other projects is funded then we believe that the labour estimates provided in this proposal are sufficient to provide the base set of functionality required for genotypic data management for our target user group.

We are not aware of any equivalent open source software applications available nationally or internationally and we are unaware of any viable alternatives for researchers that do not have the required funding or expertise to purchase commercial solutions or develop their own solutions. Smaller studies are often able to operate using simple tools, such as Microsoft Access or Microsoft Excel, but there is no doubt that the research being undertaken by these studies is being compromised by the lack of suitable informatics support.

# Research Community Profile (1 page)

*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.*

Users are both clinical and basic science researchers who use clinical specimens and health information from any disease type for molecular and cellular analyses. Current users of our software encompass cancer, diabetes, cardiovascular, obesity, eye diseases, mental health and infant and child health. The platforms would also support use by those collecting specimens from clinical trial, both for academic and commercial purposes.

MORE

# Development Organisation Profile (2 pages)

*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.*

The software development will be undertaken primarily by the following organisations:

1. The University of Western Australia Centre for Genetic Epidemiology & Biostatistics. The Centre for Genetic Epidemiology and Biostatistics comprises a multi-disciplinary team of genetic statisticians, genetic epidemiologists, mathematicians, epidemiologists, bioinformaticists, molecular biologists, and social scientists committed to developing ways of investigating the determinants of complex human disease and exploring ways of using genetic information to improve human health.
2. The Ark ([www.the-ark.org.au](http://www.the-ark.org.au)) project team at the Centre for Genetic Epidemiology & Biostistics, UWA, currently consists of three full-time professional software developers and one part time project manager. The Ark team has been developing and supporting informatics software to support the Centre for Genetic Epidemiology and Biostatistics and collaborating institutions since 2004. Software tools developed by the team has been in production since 2005 and currently provides support to a number of research groups, including:
   * The Western Australian Research Tissue Network at St John of Health Care
   * The Busselton Healthy Aging Study (The Busselton Foundation)
   * The Western Australian DNA Bank (which in turns supports over 40 studies using The Ark software)
   * The National Breast Cancer Foundation
   * The Western Australian Institute for Medical Research
   * The Lions Eye Institute

The earlier software tools were developed using a mix of C, Oracle 4GL and Java development languages. Two years ago The Ark project was initiated to develop the Java-based research tools that form the basis of this bid.

All software development has always and continues to be the result of close collaboration with a number of key research groups. This helps ensure that the software is always fit for purpose and is always tested using real data by future users of the applications before production deployment. Continual feedback is encouraged and managed by providing most researchers with online access to Jira ([www.atlassian.com/en\_AU/software/jira/](http://www.atlassian.com/en_AU/software/jira/)), the web-based issue tracking system used by The Ark team. Researchers are able to submit issues and enhancements online and track their progress online. All issues are reviewed weekly and prioritised. Authorised researchers have access to view the current status of any of the issues associated with The Ark Jira instance.

All researchers are also able to access The Ark’s wiki ([wiki.genepi.org.au](http://www.wiki.genepi.org.au)) to view ongoing analysis, design and user documentation.

Primary responsibility for supporting a The Ark module is assigned to a specific developer but all developers have sufficient familiarity with all of the software to provide backup support. Care has also been taken to ensure that a shared set of design patterns is used by all of the developers.

The Ark team is currently funded by UWA and a degree of cost recovery but ongoing funding is not sufficient to continue to develop the software past the end of 2011.

1. The Ark project team at the Centre for MEGA Epidemiology, Melbourne University, consists of one developer with responsibility for contributing to the application development as well for providing technical and implementation support to Eastern States studies.
2. OBiBa ([www.obiba.org](http://www.obiba.org)) is an international software development project based at at McGill University, Montreal, Canada. OBiBa is committed to building a full suite of open source software for biobanks. It is comprised of several independent and self-funded teams around the world, each of which is producing stand-alone applications that support particular biobank activities. The applications can be customized and integrated to create a complete biobank information management system.

OBiBa offers a collaborative infrastructure to its teams and to other developers who may wish to join the OBiBa community. The infrastructure for developers includes integrated tools for documentation, issue tracking, deployment, and project management. The Ark project is a member of OBiBa. Paul White is one of the Principal Investigators of the OBiBa project.

1. The Australian Twin Registry at Melbourne University has been developing software to support the operation of The Australian Twin Registry for several years. The ATR’s development and operational staff have and will continue to contribute to the specification and testing of The Ark software.

Please refer to the attached letters of support in Appendix A

# Operational Organisation Profile (2 pages)

*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.*

**The Centre for Genetic Epidemiology & Biostatistics at the University of Western Australia.**

The Ark project team at the Centre has been supporting hosted bioinformatics solutions for Australian human medical researchers since 2005. The team has been and continues to support researchers from public research institutions and private practice. Researchers with the following affiliations are currently being supported by the group:

* The University of Western Australia
* The Western Australian Institute for Medical Research
* Royal Perth Hospital
* Sir Charles Gairdner Hospital
* Lions Eye Institute
* The University of Melbourne
* The Murdoch Childrens Research Institute
* St John of God Healthcare
* The Western Australian DNA Bank
* Peter McCallum Research Institute
* The National Breast Cancer Foundation

**The Centre for MEGA Epidemiology, the University of Melbourne**

Currently the The Ark programmer in Melbourne supports a number of researchers, including one research group that is using The Ark’s first production release.

**OBiBa**

The OBiBA project team at McGill University in Montreal, Canada provides operational support to a number of large Canadian research studies. The OBiBA technical resources will be available to provide technical support for this project as required.

**Specific Software Support Experience**

The 3 teams currently support a mix of open-source and proprietary software used for the following purposes:

* Laboratory specimen management (LIMS)
* Participant Management
* Phenotypic Data Management
* Genotypic Data Management
* Pedigree data management and visualisation
* Paper-based and electronic questionnaire data collection

All teams utilise the same support tools and similar processes. Please refer to section 18 - Operations and User Support for a list of the support tools currently in use.

Please refer to the attached letters of support in Appendix A

# Other Participants (0.5 pages)

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

St John of God Health Care (Private partner)

Clinical Oncological Society of Australia- cooperative cancer trials groups

WAIMR

DoHWA – Di Rosman

Curtin – James Semmens

Busselton Health Study

Cancer Institute of New South Wales

National Breast Cancer Foundation

Foundation Health Consumers Council of WA- consumer engagement: In WA there is a strong history of consumer engagement with clinical datasets through the WA Health Data Linkage Unit. These pathways would be used to ensure that the governance of the proposed work is consistent with consumer expectations.

# Key Personnel (0.5 pages)

*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.*

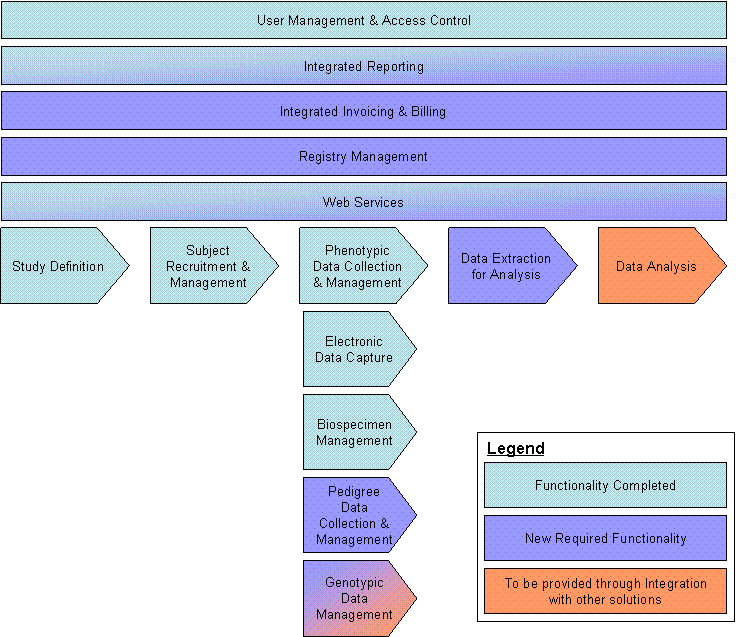
|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Organisation** | **Role** | **Availability** |
| Paul White | Centre for Genetic Epidemiology and Biostatistics, University of Western Australia | Project Manager | 40% |
| Professor Eric Moses | Centre for Genetic Epidemiology and Biostatistics, University of Western Australia | Project Sponsor | 5% |
| Dr Nik Zeps | St John of God Health Care | Ethics, pathology based -biobanking, clinical trial biobanking | 5% |
| Christopher Ellis | Centre for Genetic Epidemiology and Biostatistics, University of Western Australia | Technical Architect and Senior Developer | 100% |
| Programmer (individual to be determined) | Centre for Genetic Epidemiology and Biostatistics, University of Western Australia | Software Developer | 100% |
| Adrian Bickerstaffe | Centre for MEGA Epidemiology, Melbourne University | Software Developer | 50% |
| Philippe Laflamme | OBiBa Project | Technical Architect | 5% |
| Andrew Mews | St John of God Health Care | Business Analyst | 20% |
| Kelly Aujard | Centre for MEGA Epidemiology, Melbourne University | Business Analyst | 5% |
| Rhonda DeSouza | Clinical Oncological Society of Australia (COSA) | Business Analyst | 25% |
| Lisa Spalding | St John of God Health Care | Business Analyst | 10% |
| Support Programmer (individual to be determined) | Centre for Genetic Epidemiology and Biostatistics, University of Western Australia | Software Developer | 100% |

None of the above individuals has any commitments that will conflict with his or her availability during the project period in any major way.

# Infrastructure (2 pages)

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

The logical functional diagram below illustrates the functionality that will be delivered. The shading on the diagram indicates which functionality has already been developed, which needs enhancements and which is intended to be provided through integration with third party software.



**User Management & Access Control (In production)**

A single instance of the software already supports multiple users working with multiple studies across multiple locations.

Security features incorporated into the Ark include:

* Single sign on;
* Fine grained study and user access control;
* Clear separation of identifying and de-identified data;

• Secure web access and data transfer; and

Integration with the AAF services is planned for user authentication.

**Integrated Reporting (In production with enhancements planned)**

An integrated reporting tool provides the ability for researchers to report across the different data areas supported by the system, subject to data access constraints. Extensions to this tool are planned in order to provide the ability to report across multiple research data types.

**Integrated Invoicing and Billing (Planned)**

An integrated invoicing and billing system will provide the ability to track resources that have been expended supporting an internal or external research project.

**Registry Management (Planned)**

The registry management module will allow registry and study administrators to track which other studies have been provided with access to their participants and what the status of those studies is. There are also additional reporting requirements for Registry Management.

**Study Definition (In production)**

The Study Management module provides a System Administrator with the ability to define new studies or registries and then designate a new or existing user as the administrator for that study or registry. This module also allows users with Study Administrator rights to then configure the specific study. Functions available to the Study Administrator include:

• Maintaining the details about a study or registry

• Creating new users and assigning access rights

• Defining study protocols and consent components

**Subject Recruitment & Management (In production)**

The participant Management module incorporates the functionality required for tracking prospects and participants throughout the lifetime of the project. Functionality includes:

• Management of subject demographic data

• Consent management

• Contact management

• Correspondence management

**Phenotypic Data Collection & Management (In production)**

The Ark’s phenotypic data management module provides researchers with the ability to store and manage phenotypic/clinical data without the need to engage in database design or coding. An overarching data dictionary, that can be defined through the user interface or loaded from file, ensures that all phenotypic data within the system adheres to quality standards. Data elements can be grouped together as required for storing questionnaire, biochemistry, clinical or any other type of textual data.

Data can either be bulk loaded from other sources as text files with loaded data being flagged for consistency with the defined data dictionary so that inconstant data values can easily be corrected.

**Electronic Data Capture (Undergoing final testing)**

Data entry forms can be created dynamically using the defined data dictionary, negating the need for custom web development whenever there is a need for researchers to enter clinical, questionnaire or any other type of textual data.

**Biospecimen Management (Undergoing final testing)**

The Ark incorporates a laboratory information management system (LIMS) that provides the functionality required for tracking biospecimens. The history and location of all biospecimens is maintained along with the essential data about each specimen. The data attributes collected about each specimen can be extended dynamically without the need for additional software coding.

Genotypic Data Storage

Utilities to support the management of large volumes of genotypic data are currently under development.

**Pedigree Management & Visualisation (Planned)**

The pedigree management and visualisation module will permit the capture of pedigree information by establishing the relationships between subjects as part of the contact management process or through data import. Researchers will then be able to view these relationships diagrammatically. Note that we have developed a prototype for a 3 dimensional pedigree visualisation tool that may be used as the basis for this module. Alternatively integration with an existing open-source toolset may be provided.

**Data Extraction for Analysis (Planned)**

The data extraction for analysis module will provide study administrators and researchers, where permitted, with a software wizard to define, store and execute queries for extracting data from the system in formats that can be directly imported into analysis tools, such as SAS, SPASS or Stata.

**Genotypic Data Management & Analysis**

The preferred approach to providing this functionality is to integrate The Ark tools with the capabilities that would be provided through the Genomics Virtual Laboratory, Galaxy/GDR Integration and Service Centre Data Handover projects, should they be funded. Note that we have developed a genotypic data management prototype in conjunction with the OBiBa project at McGill University in Montreal that could form a basis for this capability should the other projects not be funded. This prototype permits users to load sequence and SNP data from proprietary formats into a standard internal format that supports:

* Linkage with other data types
* Efficient storage and extraction for analysis
* A degree of in-situ statistical analysis

**Data Analysis**

The objective of this project is not to build and support data analysis tools. Rather the objective of this project is to provide those infrastructure components that are not readily available and affordable to facilitate the collection and management of medical research data.

**RESEARCH COMMUNITY NEEDS & BENEFITS**

# Target Research Community (0.5 pages)

*Identify the research communities and the expected number of users.*

All clinical and basic science researchers that use human biospecimens, health data or both will potentially obtain a benefit from this infrastructure. With an increasing need to examine the molecular basis of drug interactions all clinical trials will also have a requirement for infrastructure to enable management of specimens, clinical data and outcomes. Outside of industry such infrastructure is non existent. The Cancer Cooperative Clinical Trials Groups (CCTGs) under COSA are included here as one example of clinical trial usage that will result from this initiative and this includes 16 CCTGs and several hundred investigators.

It is therefore somewhat difficult to specify the precise number of users of this infrastructure but based upon the groups identified in the proposal there are several groups each in WA, NSW, ACT, VIC, QLD and SA that will make immediate use of this either directly or through the collaborations that will be enabled. The listed investigators comprise a research community of over 500 people and it is likely the broader utility will encompass several thousand Australian researchers. There are already several international collaborations underway and therefore the numbers will increase with those too.

# Needs and Impact (2 pages)

*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.*

*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.*

In Australia there are a diverse range of groups that use specimens linked to health data and these cover all aspects of human disease. Many of these organisations were previously supported by the NHMRC through the enabling grant scheme which provided some funding toward their everyday activities but in general very little specifically for their IT capabilities and needs. Briefly, there are three main categories of populations about which researchers require information to be managed. They are:

1. Healthy controls
2. People with diseases receiving standard care in routine healthcare settings
3. People with diseases on clinical trials

For each of these there needs to be comprehensive clinical phenotyping and at present none of this is held in a way that can be interrogated easily. There are myriad ‘databases’, many in fact being excel spreadsheets or clinical management systems that can’t be searched readily, eg pathology laboratory information systems or outpatient records. Once identified the information has to be extracted and held in a secure system to comply with ethical and legal requirements. Again, there is no standard by which this information is compiled and over the years many groups have developed their own, none of which can ‘talk’ to others as the data definitions are often different. It is impossible to describe how large this problem is in a short space but it is perhaps sufficient to state that even within one cancer centre oncologists cannot readily find out what patients were actually given so that a comparison of treatment with basic demographic information can be performed. Whilst organisations like BioGrid in Victoria have made significant advances in addressing it they do not have scope to merge with biobanks or research data like genotyping or proteomic information and are unlikely to be able to do so in the short to medium term.

Within the biobank community there are also myriad approaches to databases and each has its own difficulties and limitations. Through the Australasian Biospecimen Network we are aware of several of these and we have been able to review their pros and cons. We have also done this with groups internationally such as the Uppsala Biobank and CTRnet in Canada. Again, none of these activities have solved being able to bring together information from the clinic, biobank and laboratory in a way that researchers can easily use it. Significant investment in the IT capabilities has been made and is ongoing in each of these jurisdictions and it is clear that without similar investment in Australia we have no possibility of remaining competitive in health research of this nature.

Clinical trials have become a central focus for Australia illustrated by the Clinical Trial Action Group (CTAG) report. Whilst the recommendations in that document focus on being able to make trials known to potential participants there is a major need to be able to support the trials themselves, particularly with respect to managing samples linked to patients on the trials. Many new agents have specific biological effects and it is fair to say that all new drug trials from industry or cooperative trial group have a biological sub-component. As we enter the era of targeted therapies the need to have the ability to integrate genomic and clinical data becomes more pressing. COSA have recently published a report (By Deloitte) highlighting this for the cancer trials groups. The Royal College of Pathologists Australia (RCPA) have commissioned a working group examining biobanking and the role of pathology recognising this as an important area of need.

Therefore, having a means to collect and store data in a manner in which it can then be interrogated and linked to biological specimens and data arising from their analysis will significantly enable our capabilities to be competitive in discovery, translational and clinical research.

There are also significant operational benefits to be derived from the proposed cloud-based bioinformatics tools:

* Research projects that would be too costly to conduct due to constraints on time and funding become feasible;
* Existing funding can be spent more efficiently. For example, data can be collected using web-based forms or scannable documents so researchers no longer need to manually transcribe research data from paper into electronic formats for analysis;
* Data quality can be improved by reducing transcription errors and by defining and enforcing data validation rules at the point of data collection or aggregation;
* Data is more secure, from both a loss perspective as well as from a privacy perspective;
* Collaboration between researchers on the same project is improved as geographically dispersed researchers can work securely together with a single copy of the data;
* Researchers will be able to reduce the amount of funding being requested from grant bodies while at the same time strengthening their applications;
* Management of Post grad research projects will be easier and cheaper;
* Preparation time for research papers will be reduced;
* Reporting requirements and ethics compliance will be easier to satisfy.

# Broader adoption (0.5 pages)

*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.*

As a result of having provided hosted informatics tools to the Australian human medical research community for several years we have created visibility of our objectives and capabilities across a number of Australian research institutions. Many of these organisations have expressed their frustration at the lack of availability of cost-effective informatics solutions and at the inability of available tools to provide an integrated view of their research projects. We believe that once the tools are available within the Research Cloud then uptake will be very strong. The long-term objective is to have other organisations become part of The Ark project from both a development and support perspective.

Working collaboratively with the other NeCTAR project groups, the Genomics Virtual Laboratory, Galaxy/GDR Integration and Service Centre Data Handover will also raise awareness of our capabilities.

The Ark software is designed to be extremely flexible and user-configurable. As a result the software will support a broad range of different medical research and clinical trials groups where multiple data types need to be associated with a single physical entity. The software and data linkages are written in such a way that they can readily link to electronic health record systems as they come online

The software is also applicable to other research disciplines as demonstrated by:

1. We are currently working with a commercial partner, Patrick Rose, to configure the system to support the collection of health and incident data in remote communities in regional Western Australia;
2. The UWA Crime Research Centre has approached us to regarding using The Ark software to manage crime research data.

# Value adding (1 page)

*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*](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.*

The Ark project builds upon a software development and research project initially funded by an NHMRC Enabling Grant. This predecessor project, commonly referred to as WAGER, has provided a solid foundation in terms of informing the functionality, design, development and support approaches that have been adopted by The Ark team. Through the course of the WAGER project the team also built strong relationships with a number of Australian research groups, some of which are collaborators and supporters of this grant application.

The Ark team have been working with the Western Australian Data Linkage Unit since 2005 to facilitate the linking of research data with public health data. The Ark team intends to continue this relationship with the Western Australian Data Linkage Unit and with the Australian-wide equivalent - The Population Health Research Network (PHRN).

The Ark project will integrate with existing capability platforms as follows:

* The existing The Ark LDAP authentication services will be integrated with AAF so that existing researcher credentials may be used to access The Ark software;
* Over the course of 2012 The Ark will retire its existing server hardware and migrate all of its existing applications to the Research Cloud. Note that The Ark already has access to two National Server Program virtual servers. The applications on these servers will be moved to the Research Cloud at the appropriate time;
* One of the objectives of The Ark project team is to position researchers to take advantage of the significant HPC capability currently being provided by NCI. To gain new insights from the increasingly large data sets being made available through technologies such as high throughput screening will require seamless access to high performance computing hardware and the relevant software tools. The preferred approach to achieving this is to integrate The Ark with the capabilities to be delivered by the Genomics Virtual Laboratory, Galaxy/GDR Integration and Service Centre Data Handover NeCTAR projects.
* The Ark project team currently has its own SAN for data file storage. It is planned that during the course of the NeCTAR project this will be retired and RDSI data storage will become the preferred storage medium for image and genotypic data files; and
* There is the possibility that, if the issues associated with fully de-identifying data sets can be addressed, then some of the collaborator data sets could be moved to ANDS. This is not currently within scope for this proposal.

There are a number of components of The Ark solution that could be leveraged by other research communities, including:

* The Laboratory Information Management System (LIMS) for biospecimen management; and
* The Phenotypic data management module for storing phenotypic data about an entity.**PROJECT MANAGEMENT**

# Governance (1 page)

*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.*

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

The Steering Committee will have carriage of all aspects of the design and conduct of The Ark Bioinformatics Tools project and will be accountable for assessing project performance. The Ark Manager, Paul White, will in turn report to the Steering Committee. YYY will report directly to the NeCTAR XXX.

The Steering Committee will have the following membership:

* Professor Eric Moses, Director of the Centre for Genetic Epidemiology and Biostatistics, UWA
* Dr Nik Zeps, Dr Nik Zeps, Research Manager, St John of God Health Care
* Professor John Hopper, Professor & Director (Research), Centre for Molecular, Environmental, Genetic and Analytic (MEGA) Epidemiology, Melbourne School of Population Health
* David XX, Clinical Oncological Society of Australia

Insert a diagram showing the governance structure – incl. Steering Committee

We manage our projects using the Agile methodology. A key principle of Agile is its recognition that during a project the customers can change their minds about what they want and need, and that unpredicted challenges cannot be easily addressed in a traditional predictive or planned manner. As such, Agile adopts an empirical approach — focusing on maximizing the team’s ability to deliver quickly and respond to emerging requirements.

The Project Manager will manage the project by coordinating and communicating with the rest of the team. The collaborating research groups will be kept up to date on major project issues and sprint results through email updates and Work in Progress (WIP) meetings, emails and/or phone calls.

Risk management is a key aspect of project management. Project specific risks will be identified, documented and managed. The earlier the risk can be identified, quantified and effectively contained, the more successful the risk management process will be.

Insert a list of project management deliverables.

# Project Scale (0.5 pages)

*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.*

The development scope of the project equates approximately to 2.5 software programmers for 10 months with the corresponding project management and business analyst/testing support. Subsequent to the completion of the development phase operations support will be provided by the equivalent of one full time support officer.

The total effort associated with the project is 1080 FTE days or 4.9 FTE years at a total cost of approximately $630,000. We are asking for $306,000 from NeCTAR for this project.

Resources contributed by the other participants include:

* St John of God Health Care – two part time business analysts, both with extensive experience in the biobanking and medical research area;
* The Centre for MEGA Epidemiology – one 50% software developer and one part time business analyst/software developer;
* The Clinical Oncological Society of Australia – one 25% business analyst with considerable experience in conducting clinical trials;
* The OBiBA Project – one part time technical architect;
* The University of Western Australia – infrastructure support and a cash contribution to cover non-salary expenses.

A number of other organisations have also expressed a willingness to contribute to the project by participating in the requirements analysis and testing phases of this project but these resources have not been included in the project budget.

# Project Approach (1 page)

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

The required services will be delivered using the following approach:

1. Establishment of support tools and processes for both software development and ongoing customer support. Note that many of these processes are already defined but will need to be modified to support operations within the Research Cloud. The Ark project has already established a test environment on the National Server Program infrastructure. This will need to be migrated to the Research Cloud.
2. Integration with AAF authentication services
3. Initial Research Cloud deployment. This activity will include comprehensive system, security and user acceptance testing to ensure existing software is cloud ready. This will also include deployment to a production environment in the Research.
4. Incremental development of functionality to support collaborative partner requirements. The software will be developed as a set of discrete modules that can be plugged into the overall application framework. Within each module the software will be developed in a number of sprints. Each sprint will deliver a set of working functionality that can be tested as a functional deliverable.
5. Deployment of each module in a production environment for access by the project collaborators.
6. Staged deployment to the broader research community:

Each of these sub-projects/deliverables will be developed and delivered using our standard development process whereby each project goes through a number of stages:

1. **Project Definition** - The project stakeholders and the development team document the problem to be solved and the key criteria acceptance. A product Owner is identified for each project/deliverable.
2. **Elaboration** – The documentation and tools required to manage the project are developed, including the risk management plan, project management plan, user stories/requirements definition documentation and the Product Backlog. The product backlog is a living document which changes whenever the Product Owner receives new information. The Product Backlog helps guide the team in selecting work that best reflects business priorities for each iteration of development.
3. **Development** – Application development will be done in a series of two to four week sprints. Each sprint commences with a Sprint Planning Meeting where the development team selects items for the Sprint. These items are then frozen in the Product Backlog by the Product Owner and are not changed for the duration of the Sprint. The outcome of each sprint is a set of functioning tested code.
4. **Deployment** – After the final sprint the software has been tested and documented and is ready for deployment into production. We will assist our collaborators in conducting additional acceptance testing on each module using production data that has been migrated from existing systems.

# Key Deliverables and Acceptance Criteria (1 page)

*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.*

|  |  |
| --- | --- |
| **Key Deliverable** | **Acceptance Criteria** |
| Support tools & processes | Appropriate tools and processes are documented and deployed.  Support staff is suitably trained. |
| Software integrated with AAF authentication | User creation and maintenance functionality is integrated with AAF authorisation service  User authentication is implemented through AAF interfaces |
| Initial production Research Cloud deployment | Software is available in a production environment in the Research Cloud or on National Server Program infrastructure (we currently have access to 2 NSP virtual servers)  Software has passed user acceptance using real data. |
| Integrated Invoicing & Billing | Resource utilisation for internal and external service provision can be tracked  Cost recovery processes are supported |
| Data Extraction for Analysis module | Flexible data selection for extraction  Data for chosen subjects can be extracted in specified formats  Data can be de-identified as appropriate |
| Pedigree Storage & Visualisation Module | Relationships between participants is easily captured and maintained  Pedigrees can be visualised and explored |
| Enhanced Data Linkage & Reporting capability | Data sets matching chosen criteria can be identified using criteria that span multiple data sources and types |
| Registry Management module | Access to and usage of subjects and data by third parties can be captured, controlled and monitored |
| Genotypic Data Management capability | Genotypic data can be aggregated and stored efficiently  Genotypic data can be accessed easily for analysis |

# Quality Control (1 page)

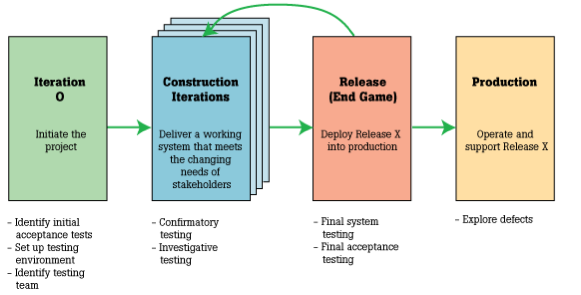
*Identify the personnel, processes and any special resources that will be required for Quality Control, Acceptance and Commissioning Testing activities on the proposed project.*

*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.*

The Ark team will follow an Agile software development approach where working software will be developed in a number of incremental sprints.



Members of The Ark project team will be responsible for developing the functional requirements and acceptance tests in conjunction with the business analysts from the collaborating organisations. Representatives from the collaborating organisations will be responsible for signing off these requirements and acceptance tests. All requirements/user stories and acceptance criteria will be documented in Confluence, the project wiki.

All issues identified, be they requests for enhancements or bugs, will be documented in Jira, the project issue tracking system which will be made available online to all project participants.

**Confirmation Testing**

Throughout each development iteration or sprint the software will undergo continual confirmation testing to ensure that it is functioning correctly and to ensure that new software doe not “break” existing functionality. This confirmation testing will be done by the business analysts on the project. The Hudson continuous integration tools will be used to ensure that as software is checked into Subversion, the source code repository, it will still compile and run.

**Usability Testing**

Visual mockups will be created for the application screens using Balsamic, a mockup tool that has been integrated with the Confluence wiki. Representatives from the collaborating organisations will be responsible for signing off these user interface designs.

**System Testing**

The business analysts will be responsible for conducting system testing for each module to ensure that the software developed meets the requirements and acceptance criteria defined.

**Security Testing**

The Ark project team will be responsible for conducting security testing in accordance with any guidelines developed by the AAF and NeCTAR for applications hosted on the Research Cloud.

**User Acceptance Testing**

User Acceptance Testing will be undertaken by our partner organisations using real production data. Prior to each cycle of User Acceptance testing we will migrate production data into a secure testing instance for these partners to test in order to ensure that both the migration scripts and the application functionality, performance, security, etc. meet the required levels.

CHANFGE CONTROL

# Risk and Issue Management (1 page)

*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.*

The key risks to the successful delivery of the proposed project include:

* Availability of appropriately skilled development resources. This risk is significantly mitigated through the ability of the project to retain development staff at competitive salaries;
* Ongoing availability of partner organisation resources. Each of the partner organisations has made a commitment to providing these resources but in the event that availability becomes an issue then alternative resources will need to be identified.
* Availability and support of infrastructure services, including RDSI, AAF and the Research Cloud. Successful implementation will be dependent on these facilities being available at the appropriate time points.

There are no open issues that need to be resolved before the project can start delivery.

There are no key questions that need to be resolved before the project can start delivery however the scope and approach to be taken to the delivery of the genotypic data management functionality will be dependent on the success of the complementary NeCTAR proposals.

**Risk Management Approach**

The Project Manager will develop a risk management plan that will detail the approach to be taken to risk identification; risk quantification; risk mitigation; and risk monitoring and control.

All project risks will be documented within a risk register and assigned a risk rating. Risk mitigation plans will then be developed for each of the major risks and then an ongoing process will be implemented to continually monitor risks to identify any changes in status, to identify outstanding actions, to remove risks that have passed and to identify new risks.

**LEVERAGING**

# Standardisation and Interoperability (0.5 pages)

*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 be incorporated by the project.*

The Ark team is developing all of the software to be as flexible as possible so that it can be used to support studies that have chosen to adopt specific ontologies or have developed their own coding standards for data storage. The ability for researchers to define their own data dictionary and data entry forms as well as to define custom fields exists in all modules. All field definition is by way of a graphical user interface.

All software development will be done according to the relevant W3C software development standards.

The Ark team and the project supporters will work collaboratively to encourage the adoption of standard sets of operating procedures and common ontologies across research projects within a given research disease discipline.

* The project team will work with the (P3G) Public Population Project in Genomics’ DataShaper group ([www.datashaper.org](http://www.datashaper.org)) to encourage the adoption of their Data Schema and Harmonization Platform for Epidemiological Research.
* The project team will collaborate with leading proponents of common SOPs, such as The Canadian Tumour Repository Network ([www.ctrnet.ca](http://www.ctrnet.ca)).

The project team will work together with other funded NeCTAR funded project teams that are developing genotypic data management and analysis capabilities. A common set of web services will be defined so that our software is able to leverage the capabilities being developed by the Genomics Virtual Laboratory, Galaxy/GDR Integration and Service Centre Data Handover projects.

**FINANCIAL**

# Budget Breakdown (1 page)

*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.*

| **Milestone** | **Associated Deliverable Start Date** | **Milestone Date** | **EIF Funding** | **Co-investment Funding** |
| --- | --- | --- | --- | --- |
| **Established Support Tools & Processes** | 1-Mar-12 | 15-Mar-12 | **$8,333** | **$487** |
| **Funding Milestone 1** | **1-Mar-12** | **31-Mar-12** | **$68,856** |  |
| **Integrated existing application with AAF Authentication Services** | 15-Mar-12 | 30-Apr-12 | **$21,527** | **$3,442** |
| **Integrated Invoicing & Billing** | 1-Mar-12 | 30-Jun-12 | **$14,605** | **$40,163** |
| **Initial Production Research Cloud Deployed** | 1-Mar-12 | 30-Jun-12 | **$59,180** | **$11,937** |
| **Funding Milestone 2** | **1-Apr-12** | **30-Jun-12** | **$68,856** |  |
| **Implemented Data Extraction for Analysis Module** | 15-Jun-12 | 15-Aug-12 | **$6,491** | **$1,623** |
| **Implemented Pedigree Storage & Visualisation Module** | 1-May-12 | 30-Aug-12 | **$62,475** | **$12,373** |
| **Funding Milestone 3** | **1-Jul-12** | **30-Sep-12** | **$68,856** |  |
| **Enhanced Data Linkage & Reporting Module** | 1-Sep-12 | 30-Nov-12 | **$47,775** | **$9,462** |
| **Implemented Registry Management Module** | 1-Jul-12 | 30-Nov-12 | **$17,039** | **$37,335** |
| **Integrated Genotypic Data Management Capability** | 15-Aug-12 | 31-Dec-12 | **$68,605** | **$19,708** |
| **Funding Milestone 4** | **1-Oct-12** | **31-Dec-12** | **$68,856** |  |
| **Funding Milestone 5 - 10% Retention Payment** |  |  | **$30,605** |  |
| System Support | 1-Jan-13 | 30-Jun-14 |  | $189,000 |
| Travel Budget (UWA Centre for Genetic Epidemiology & Biostatistics) |  |  |  | $20,000 |
| Infrastructure Support (UWA Centre for Genetic Epidemiology & Biostatistics) |  |  |  | $0 |
| **Total EIF Request** |  |  | **$306,029** |  |
| **Total co-investment** |  |  |  | **$345,528** |
| **Total Project Budget** |  |  |  | **$631,557** |

*Please refer to Appendix 1 - Resource breakdown by Milestone – for a detailed breakdown of allocated resource costs per milestone and deliverable.*

The system support phase of this project for 2013/2014 will be funded primarily through cost recovery. The project team has utilised a cost-recovery model for its hosted services for the last 12 months. Users of the existing proprietary hoisted services have been willing to pay between $3,000 and $25,000 per year, depending on the breadth of functionality used and the degree of support required. Cost recovery so far this year has been in the order of $70,000 for hosting support.

Access to the open-source software as it is deployed will not be restricted only to the collaborating organisations. Any researcher may gain access to the source code and deploy it on their choice of infrastructure. Given sufficient demand, new staff may also be allocated at the UWA Centre for Genetic Epidemiology and Biostatistics to support additional instances of the software in the Research Cloud.

**SERVICES AND SUPPORT**

# Service Levels (0.5 pages)

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

The Ark service desk personnel will primarily be located at the Centre for Genetic Epidemiology at the University of Western Australia with additional support being provided out of the Centre for MEGA Epidemiology at the University of Melbourne. This will ensure that support personnel will be available from 9am to 5pm Monday to Friday for all Australian locations.

All The Ark users will have access to an instance of the Jira issue tracking software application hosted in the Research Cloud to enable them to log and then follow the progress of issues, including bugs, enhancements and requests for assistance with tasks like project configuration and data migration.

Customer support will primarily be provided through the Jira issue management software with response times by support staff of less than 24 hours on average.

The service levels provided by The Ark software will be dependant on the service levels being offered by the NeCTAR Research Cloud as it is intended that all The Ark software will be hosted in the cloud. Independent of the Research Cloud availability, The Ark will endeavour to provide access to the application software at a 95% level between the hours of 9am EST time to 5pm WA time. All major software upgrades will be conducted after 3pm WA time to ensure maximum availability for Eastern States users.

All The Ark documentation is currently and will continue to be hosted on The Ark’s Confluence wiki. This is currently hosted on our own servers but will be migrated to the Research Cloud in the near future.

Training will be conducted using a number of mechanisms:

* Face-to-face training will be conducted by The Ark staff from either Melbourne or Perth. Where travel is required then these costs will be borne by the research institution receiving training.
* Online video presentations. The intent is to develop a number of online video training sessions that will be accessible from The Ark website.
* How-to documentation will be developed that provides a step-by-step approach to configuring and using The Ark tools.

# Operations and User Support (1 page)

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

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

The Centre for Genetic Epidemiology and Biostatistics at the University of Western Australia will operate the software tools implemented by The Ark project. The support mechanisms provided to users will include:

* Online support by providing all users with access to the Jira issue management tracking tools;
* Telephone support as appropriate;
* Access to online documentation through The Ark’s wiki; and
* Access to answers to frequently asked questions on The Ark’s wiki.

Significant face-to-face support will also be provided to the project collaborators during the development, implementation and support stages of the project.

MORE DETAIL

# Sustainability (0.5 pages)

*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.*

During the development stages of the project under the NeCTAR grant the project infrastructure will be funded by a mix of key collaborator in-kind and cash contributions and a fee-for service model. A fee-for service model is already in use with 12 research projects already paying for access to the pre-existing non open-source version of The Ark’s software. It is proposed that over the course of 2012 these projects will be migrated to the open-source version of The Ark software. The current fee for service model has the following features:

* Hosting and support charges are on a per-module basis. Researchers have access to and only pay for the modules that they require;
* Hosting and support charges are independent of the size of the research project;
* Configuration and training is charged on a per module basis;
* A certain amount of disk storage is available for each study. Additional storage attracts a per Terabyte charge;
* Fees are for access within a calendar year – charged pro rata;
* All client-specific work is charged on a per hourly basis;

Changes to the existing business model to support the open-source software include:

* Researchers may choose to host their own instance of the software and may elect to enter into a technical support contract as long as they are running a supported, non-modified version of the software;

Subsequent to the completion of the development phase of the NeCTAR project the operational costs will be funded through the fee-for-service model.

The intent is to have the production and test instances of the software hosted in the Research Cloud. Costs for this are as yet unknown but if this does incur a charge then this will be passed onto the organisations using the software on a pro-rata basis.

Currently if new The Ark functionality is requested by a specific client then the work is typically only undertaken if will contribute to the main branch of the software, even if the client is willing to pay for the development. During the course of the NeCTAR project the new functionality will be driven by the requirements of the key project collaborators.

# IP, Licensing and Access (0.5 pages)

*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.*

All software development will be undertaken under the GPLv3 open source software license. The GNU General Public license is a copyleft license for general use, which means that derived works can only be distributed under the same license terms. Under this philosophy, the GPL grants the recipients of a computer program the rights of the free software definition and uses copyleft to ensure the freedoms are preserved, even when the work is changed or added to.

There will be no other Intellectual Property or licensing restrictions relevant to the services that will be delivered.

There will be no restrictions on access to the services that will be delivered.

Software licenses used for software development will include:

* Macintosh Operating System
* Windows Operating System
* Centos Linux Operating System (Local & NeCTAR Research Cloud)
* Navicat Data Modelling Tool
* MySQL Database (open source license)
* Atlassian Jira issue tracking and project management (open source license)
* Atlassian Confluence wiki (open source license)
* Subversion source code management
* Hudson Continuous Integration
* Microsoft Project
* OpenLDAP

# Communications and Engagement (0.5 pages)

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

An Agile software development approach will be used for the project. Software will be developed in short iterations, typically 2-4 weeks. Each iteration will involve a team, including stakeholder representatives, working through a full software development cycle including planning, requirements analysis, design, coding, unit testing, and system testing. At the end of each iteration a working product will be demonstrated to stakeholders. This will minimise overall risk and allow the project to adapt to changes quickly. This approach will provide the project customers the opportunity to provide timely feedback that can rapidly be integrated into the software development process.

Subsequent to each sprint there will be a Sprint Review Meeting during which feedback will be solicited and discussed to determine product and process deficiencies and improvements.

# Constraints and Dependencies (0.5 pages)

*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.*

No expenditure, scalability or performance constraints have been identified.

The successful completion of the project is dependent on the availability and performance of other national infrastructure elements, including:

* AAF Authentication services
* RDSI
* The Research Cloud
* HPC infrastructure and services

The approach and scope of the solution to be implemented for genotypic data management, analysis and HPC integration will be dependent on the funding outcome of the Genomics Virtual Laboratory, Galaxy/GDR Integration and Service Centre Data Handover projects.

# 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.

## 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) | | |

## 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) | | |

# Milestone and Funding Milestone Template

## 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 (%) |
| Centre for Genetic Epidemiology & Biostatistics at the University of Western Australia | 100% |
|  |  |
|  |  |
|  |  |

## 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.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **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 |  | 31 Jan 2012 |  |  |  |  |  |
| 2 |  | Project Initiation complete | *Communications plan prepared and sent to NeCTAR (Signed contract + two months).* | 31Mar 2012 |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  |
| 8 |  |  |  |  |  |  |  |  |  |
| 9 |  |  |  |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |  |  |  |
| 11 | Yes | Final Admin Closure | *Post-implementation Review (PIR) conducted and sent to NeCTAR.*  *Practical Completion Certificate accepted by NeCTAR.* | 30 Sep 2013 | 100  (last ten percent) |  |  |  |  |
| 12 |  | Operations to June 2014 | *Service Levels met and reported to NeCTAR as defined.* |  |  |  |  |  |  |

# Appendix 1 - Resource breakdown by Milestone

| **Milestone** | **Associated Deliverable Start Date** | **Milestone Date** | **Assigned Resources** | **FTE** | **Cost** | **EIF Funding** | **Co-investment Funding** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Established Support Tools & Processes** | 1-Mar-12 | 15-Mar-12 | Paul White | 0.15 | $1,460 | $1,460 |  |
|  |  |  | Programmer | 1 | $6,873 | $6,873 |  |
|  |  |  | Nik Zeps | 0.05 | $487 |  | $487 |
| **Milestone Totals** |  |  |  |  |  | **$8,333** | **$487** |
| **Funding Milestone 1** | **1-Mar-12** | **31-Mar-12** |  |  |  | **$68,856** |  |
| **Integrated existing application with AAF Authentication Services** | 15-Mar-12 | 30-Apr-12 | Paul White | 0.15 | $3,773 | $3,773 |  |
|  |  |  | Nik Zeps | 0.05 | $1,258 |  | $1,258 |
|  |  |  | Philippe Laflamme | 0.1 | $2,184 |  | $2,184 |
|  |  |  | Programmer | 1 | $17,755 | $17,755 |  |
| **Milestone Totals** |  |  |  |  |  | **$21,527** | **$3,442** |
| **Integrated Invoicing & Billing** | 1-Mar-12 | 30-Jun-12 | Paul White | 0.2 | $14,605 | $14,605 |  |
|  |  |  | Nik Zeps | 0.05 | $3,651 |  | $3,651 |
|  |  |  | Adrian Bickerstaffe | 0.5 | $25,773 |  | $25,773 |
|  |  |  | Andrew Mews | 0.25 | $10,739 |  | $10,739 |
|  |  |  |  |  |  | **$14,605** | **$40,163** |
| **Initial Production Research Cloud Deployed** | 1-Mar-12 | 30-Jun-12 | Paul White | 0.15 | $9,980 | $9,980 |  |
|  |  |  | Nik Zeps | 0.05 | $3,327 |  | $3,327 |
|  |  |  | Chris Ellis | 1 | $49,200 | $49,200 |  |
|  |  |  | Andrew Mews | 0.1 | $3,914 |  | $3,914 |
|  |  |  | Lisa Spalding | 0.1 | $4,696 |  | $4,696 |
| **Milestone Totals** |  |  |  |  |  | **$59,180** | **$11,937** |
| **Funding Milestone 2** | **1-Apr-12** | **30-Jun-12** |  |  |  | **$68,856** |  |
| **Implemented Data Extraction for Analysis Module** | 15-Jun-12 | 15-Aug-12 | Paul White | 0.2 | $6,491 | $6,491 |  |
|  |  |  | Nik Zeps | 0.05 | $1,623 |  | $1,623 |
|  |  |  | Chris Ellis | 1 | $22,909 | $22,909 |  |
| **Milestone Totals** |  |  |  |  |  | **$6,491** | **$1,623** |
| **Implemented Pedigree Storage & Visualisation Module** | 1-May-12 | 30-Aug-12 | Paul White | 0.2 | $13,793 | $13,793 |  |
|  |  |  | Nik Zeps | 0.05 | $3,448 |  | $3,448 |
|  |  |  | Andrew Mews | 0.1 | $4,057 |  | $4,057 |
|  |  |  | Rhonda DeSouza | 0.1 | $4,868 |  | $4,868 |
|  |  |  | Programmer | 1 | $48,682 | $48,682 |  |
| **Milestone Totals** |  |  |  |  |  | **$62,475** | **$12,373** |
| **Funding Milestone 3** | **1-Jul-12** | **30-Sep-12** |  |  |  | **$68,856** |  |
| **Enhanced Data Linkage & Reporting Module** | 1-Sep-12 | 30-Nov-12 | Paul White | 0.2 | $10,548 | $10,548 |  |
|  |  |  | Nik Zeps | 0.05 | $2,637 |  | $2,637 |
|  |  |  | Andrew Mews | 0.1 | $3,102 |  | $3,102 |
|  |  |  | Rhonda DeSouza | 0.1 | $3,723 |  | $3,723 |
|  |  |  | Programmer | 1 | $37,227 | $37,227 |  |
| **Milestone Totals** |  |  |  |  |  | **$47,775** | **$9,462** |
| **Implemented Registry Management Module** | 1-Jul-12 | 30-Nov-12 | Paul White | 0.2 | $17,039 | $17,039 |  |
|  |  |  | Nik Zeps | 0.05 | $4,260 |  | $4,260 |
|  |  |  | Kelly Aujard | 0.05 | $3,007 |  | $3,007 |
|  |  |  | Adrian Bickerstaffe | 0.5 | $30,068 |  | $30,068 |
| **Milestone Totals** |  |  |  |  |  | **$17,039** | **$37,335** |
| **Integrated Genotypic Data Management Capability** | 15-Aug-12 | 31-Dec-12 | Paul White | 0.2 | $14,605 | $14,605 |  |
|  |  |  | Nik Zeps | 0.05 | $3,651 |  | $3,651 |
|  |  |  | Philippe Laflamme | 0.05 | $3,170 |  | $3,170 |
|  |  |  | Chris Ellis | 1 | $54,000 | $54,000 |  |
|  |  |  | Rhonda DeSouza | 0.25 | $12,886 |  | $12,886 |
| **Milestone Totals** |  |  |  |  |  | **$68,605** | **$19,708** |
| **Funding Milestone 4** | **1-Oct-12** | **31-Dec-12** |  |  |  | **$68,856** |  |
| **Funding Milestone 5 - 10% Retention Payment** |  |  |  |  |  | **$30,605** |  |
| System Support | 1-Jan-13 | 30-Jun-14 | Support Analyst | 1 | $189,000 |  | $189,000 |
|  |  |  |  |  |  |  |  |
| EIF Salary funds requested |  |  |  |  |  | $306,029 |  |
| Salary co-investment |  |  |  |  |  |  | $325,528 |
| Travel Budget (UWA Centre for Genetic Epidemiology & Biostatistics) |  |  |  |  |  |  | $20,000 |
| Infrastructure Support (UWA Centre for Genetic Epidemiology & Biostatistics) |  |  |  |  |  |  | $0 |
|  |  |  |  |  |  |  |  |
| **Total EIF Request** |  |  |  |  |  | **$306,029** |  |
| **Total co-investment** |  |  |  |  |  |  | **$345,528** |
|  |  |  |  |  |  |  |  |
| **Total Project Budget** |  |  |  |  |  |  | **$631,557** |