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

**Project #:** 2179

# 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 |

# 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 |
| University of Western Australia Centre for Genetic Epidemiology & Biostatistics | Perth, WA | Lead Organisation |
| St John of God Healthcare | Perth, WA | Collaborating research organisation |
| The Clinical Oncological Society of Australia (COSA) | Sydney, NSW | Collaborating research organisation |
| The Centre for Mega Epidemiology at Melbourne University | Melbourne, Vic | Collaborating research organisation & co-development organisation |
| OBiBa | Montreal, Canada | Co-development organisation |

# Compliance Statement

## Proposed Sub-Contract Compliance

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

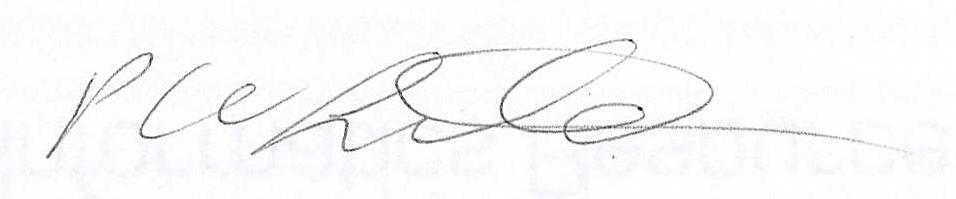
☑ **No** There are no departures from the terms and conditions (i.e. Full Compliance)

**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.

Paul White – Manager Informatics Systems Development, UWA Centre for Genetic Epidemiology and Biostatistics



Signature of authorised person making the statement Name and role (printed) Date: 02 November, 2011

## RFP Compliance

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

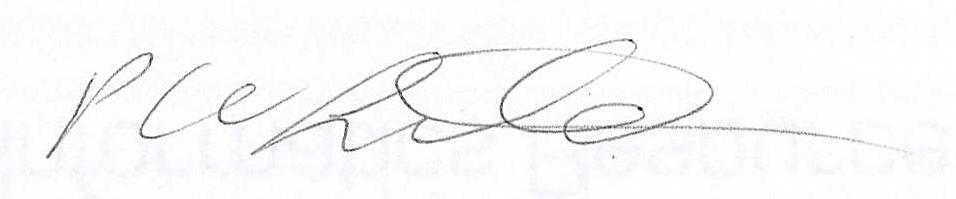
☑ **No** There are no departures from the terms and conditions (i.e. Full Compliance)

**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.

Paul White – Manager Informatics Systems Development, UWA Centre for Genetic Epidemiology and Biostatistics



Signature of authorised person making the statement Name and role (printed) Date: 02 November, 2011

## 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

|  |  |
| --- | --- |
| DISCIPLINE/FOR Code | Weight (percent) |
| Bioinformatics Software / 080301 | 70% |
| Oncology and Carcinogenesis / 1112 | 30% |
|  |  |
|  |  |
|  |  |
|  | 100% |

# Response

**SUMMARY**

# Program and Proposal Title

This proposal is addressing the eResearch Tools program.

**Short Title:** Cloud-based Bioinformatics Tools

# Executive Summary

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, for example due to new 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 analyses 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 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 using a Laboratory Information Management System(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 the 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 (family) data 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 these project be funded. Note that we have developed a genotypic data management prototype in conjunction with researchers and developers from the international OBiBa project run from 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 are not 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

**The University of Western Australia**

The Faculty of Medicine, Dentistry and Health Sciences at UWA is a research-intensive faculty, providing high-quality educational and research programs for the spectrum of medical, dental and health science disciplines.

**St John of God Healthcare**

St John of God Health Care is committed to research across a wide range of medical, surgical and mental health fields with the aim of improving patient outcomes. St John of God Healthcare operates hospitals, home nursing, pathology and Social Outreach and Advocacy services in Australia, New Zealand and the Asia-Pacific region.

**The Clinical Oncological Society of Australia (COSA)**

The Clinical Oncological Society of Australia (COSA) is the peak national body representing multidisciplinary health professionals whose work encompasses cancer control and care. COSA members are doctors, nurses, scientists and all allied health professionals involved in the clinical care of cancer patients. COSA is affiliated with and provides medical and scientific advice to [Cancer Council Australia](http://www.cancer.org.au). COSA aims to improve the care of Australians affected by cancer. In order to improve cancer care and control in Australia COSA seeks to:

* understand and provide for the professional needs of its multidisciplinary membership;
* promote and facilitate research across the spectrum of cancer care;
* promote and provide multidisciplinary and interdisciplinary education;
* contribute to, and advocate for, national issues surrounding cancer care policy in Australia;
* enhance the quality of care.

**The Centre for Mega Epidemiology at Melbourne University**

The Centre for Mega Epidemiology at Melbourne University oversees a number of longitudinal studies. The Australian Twin Registry; the Australian Breast Cancer Family Study; the Australasian Colorectal Cancer Family Study; and The Australian Longitudinal Study of Male Health have a need to manage their participant pools and to continually update data from regular and systematic follow-up studies. These data need to be securely accessible to multiple researchers across the world and there is a need for researchers to be able to return new data. Additionally there is a need to be able to track the use of data and biospecimens sent to other researchers.

JOHN – MORE ON INTERNATIONAL SPREAD PLEASE

Current and future users of the software include 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 disease, obesity, eye diseases, mental health and infant and child health. The platforms would also support use by those collecting specimens from clinical trials, both for academic and commercial purposes. The aim is to integrate clinical phenotypic data with biological data, in particular that arising from next generation sequencing technologies. The integration of these elements allows a better comprehension of environmental risk factors and how they interact with underlying genetic variations/mutations. The community that will utilise this infrastructure ranges from individual researchers right through to multinational teams bringing myriad data sets together. As such the geographical location ranges broadly depending on the project. Our intent is to enable all investigators regardless of the size of their project or their location.

# Development Organisation Profile

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

1. The University of Western Australia Centre for Genetic Epidemiology and 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 and 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 have 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 God Health Care
   * The Busselton Healthy Ageing 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
   * Lifepool (BreastScreen Victoria Cohort Study)

The earlier software tools were developed using a mix of C, Oracle 4GL and Java development languages. The Ark project was initiated two years ago to develop the Java-based research tools that form the basis of this bid, specifically because it was clear that there was demand for this functionality, both nationally and internationally, and that many researchers were not comfortable committing to a solution based on proprietary technology platforms.

All software development has 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 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 member at the Centre for MEGA Epidemiology, University of Melbourne, is responsible for contributing to the application development as well for providing technical and implementation support to local studies. This developer is responsible for the development of the system’s registry capabilities.
2. OBiBa ([www.obiba.org](http://www.obiba.org)) is an international software development project based 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, The Ark Manager, is one of the Principal Investigators of the OBiBa project.

1. St John of God Health Care is committed to research across a wide range of medical, surgical and mental health fields with the aim of improving patient outcomes. St John of God Healthcare have committed experienced research and biobanking resources to the project in analysis and testing roles.
2. The Clinical Oncological Society of Australia (COSA) is the peak national body representing multidisciplinary health professionals whose work encompasses cancer control and care. COSA members are doctors, nurses, scientists and all allied health professionals involved in the clinical care of cancer patients. COSA is affiliated with and provides medical and scientific advice to Cancer Council Australia. COSA have committed an experienced research and biobanking resource to the project in an analysis and testing role.
3. The Australian Twin Registry (ATR), run from the University of Melbourne, has been developing software to support the operation of its resource 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, especially with respect to its Registry functions.

Please refer to the attached letters of support in Appendix 2

# Operational Organisation Profile

**The Centre for Genetic Epidemiology and 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 Ark programmer in Melbourne supports a number of researchers, including one research group (Lifepool) 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 2

# Other Participants

**The Western Australian Institute for Medical Research (WAIMR)** hosts a number of researchers that have expressed a strong desire to continue to be involved in the ongoing development of the software.

**The WA Department of Health (DoHWA) and The Population Health Research Network (PHRN)** will need to be involved to ensure consistency with existing and future state and national health data linkage infrastructure.

**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.

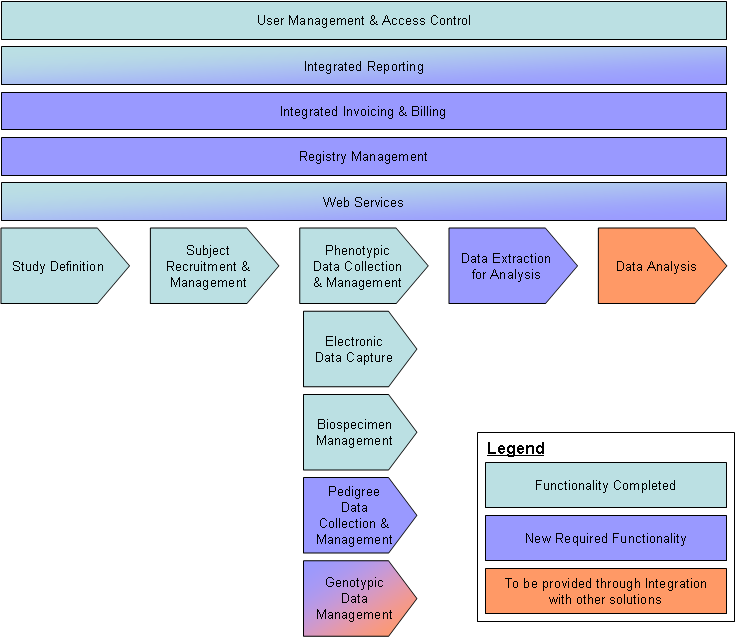
# Key Personnel

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Organisation** | **Role** | **Availability** |
| Paul White | Centre for Genetic Epidemiology and Biostatistics, University of Western Australia | Project Manager | 40% |
| Winthrop 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% |
| Professor John Hopper | Centre for MEGA Epidemiology, Melbourne University; Director, Australian Twin Registry; PI of Australian Breast Cancer Family Registry and Australasian Colorectal Cancer Family Registry | Registry Management; Promotion to other researchers and Registries | 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, University of Melbourne | Software Developer | 50% |
| Philippe Laflamme | OBiBa Project, McGill University, Montreal Canada | Technical Architect | 5% |
| Andrew Mews | St John of God Health Care | Business Analyst | 20% |
| Kelly Aujard | Centre for MEGA Epidemiology, University of Melbourne | 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 Analyst (individual to be determined) | Centre for Genetic Epidemiology and Biostatistics, University of Western Australia | Software Support | 50% |
| Support Analyst (individual to be determined) | Centre for MEGA Epidemiology, University of Melbourne | Software Support | 50% |

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

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 has the capability to support 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.

**Pedigree Management & Visualisation (Planned)**

The pedigree management and visualisation module will permit the capture of pedigree (family) information (relationships and data) 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

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 such 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 already several groups in each of 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

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 its Enabling Grant scheme which provided some funding toward their everyday core activities but in general very little specifically for their IT capabilities and needs. Briefly, there are three main categories of samples about which researchers require information to be managed. They are:

1. Healthy controls;
2. People with diseases receiving standard care in routine healthcare settings; and
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 a myriad of different ‘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 data management 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 systems with little or no regard to being able to ‘talk’ to others by using standard data definitions. 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 this issue, they do not have the 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 a myriad of 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 groups 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.

At this moment in time it is impossible to track samples linked to trials in an effective and efficient manner. Each trial independently manages information concerning pathology practices that they liaise with for blocks and blood samples. An integrated central hub that can facilitate this will save thousands of person hours each year.

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 are 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 graduate research projects will be easier and cheaper;
* Preparation time for research papers will be reduced; and
* Reporting requirements and ethics compliance will be easier to satisfy.

# Broader adoption

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 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.

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 regarding using The Ark software to manage crime research data.

# Value adding

The project is closely aligned The Promoting and Maintaining Good Health National Research Priority (NRP).

The Ark project builds upon a software development and research project initially funded by a 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 proposal.

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 ómics’ technologies 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

The project 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.

The Steering Committee will have the following membership:

* Winthrop Professor Eric Moses, Director of the Centre for Genetic Epidemiology and Biostatistics, UWA
* Dr Nik Zeps, Research Group Leader, 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 Goldstein/John Zalcberg, Clinical Oncological Society of Australia



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.

Key Project Management deliverables include:

* Communications Plan
* Project Definition Statements (Microsoft Word)
* Project Budget (Microsoft Excel)
* Project Status Reports (Microsoft Project)
* Project Management Plans (Microsoft Project)
* Project Backlog (Jira issue management software)
* Project Burn-down Charts (GreenHopper Agile software development tools)
* Issue management (Jira issue management software)
* Request for Change
* Risk Management Plan (Microsoft Word)

# Project Scale

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 (20% and 10%) business analysts, both with extensive experience in the biobanking and medical research area. One 50% support officer for the operational phase of the project (01 Jan 2013 - 30 June 2014)
* The Centre for MEGA Epidemiology – one 50% software developer and one part time (5%)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 (5%) 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

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:

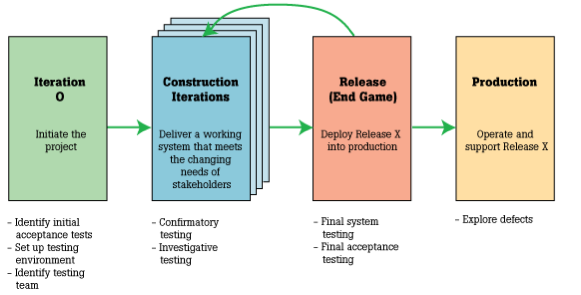
1. **Project Definition** - The project stakeholders and the development team document the problem to be solved and the key criteria acceptance. A product Owner is identified for each project/deliverable.
2. **Elaboration** – The documentation and tools required to manage the project are developed, including 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

|  |  |
| --- | --- |
| **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

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.

# Risk and Issue Management

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

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

| **Milestone** | **Associated Deliverable Start Date** | **Milestone Date** | **EIF Funding** | **Co-investment Funding** |
| --- | --- | --- | --- | --- |
| **Sub-contract signed** |  | 31-Jan-12 |  |  |
| **Funding Milestone 1** |  | **31-Mar-12** | **$55,000** |  |
| **Established Support Tools & Processes** | 1-Mar-12 | 15-Mar-12 | **$8,333** | **$487** |
| **Funding Milestone 2** |  | **31-Mar-12** | **$55,000** |  |
| **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 3** |  | **30-Jun-12** | **$55,000** |  |
| **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 4** |  | **30-Sep-12** | **$55,000** |  |
| **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 5** |  | **31-Dec-12** | **$55,000** |  |
| **Funding Milestone 6 - Final Admin Closure** |  | **30Sept-13** | **$31,000** |  |
| **System Support** | 1-Jan-13 | 30-Jun-14 |  | **$189,000** |
| Travel Budget (UWA Centre for Genetic Epidemiology and Biostatistics) |  |  |  | $10,000 |
| Other Expenses (UWA Research Matching Funds) |  |  |  | $10,000 |
| **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 by the UWA Centre for Genetic Epidemiology and Biostatistics and the project collaborators. The Centre for MEGA Epidemiology has committed a support resource at 05 FTE for the duration of the support period. St John of God Healthcare has committed to cover the salary costs for the UWA support resource at 0.5 FTE for the duration of the support period.

Supplementary income is likely to be derived through cost recovery for other research projects. The project team has utilised a partial cost-recovery model for its hosted services for the last 12 months. Users of the existing proprietary hosted 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. Additional resources, outside the scope of this proposal, will be responsible for supporting researchers from non-collaborating organisations.

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

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 from 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 WST time. All major software upgrades will be conducted after 3pm WST 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.

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

The Centre for Genetic Epidemiology and Biostatistics at the University of Western Australia and the Centre for MEGA Epidemiology at Melbourne University will support 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.

# Sustainability

Support for the project infrastructure during the operational stage through until 30 June 2014 will be funded by the project collaborators through the contribution of in-kind labour and cash.

Subsequent to 30 June, 2014 the infrastructure will continue to be operated through a cost recovery model.

Note that a cost recovery model is already in use at the UWA Centre for Genetic Epidemiology and Biostatistics for hosted software support. The current cost recovery 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;
* Fees are for access within a calendar year – charged pro rata;
* All client-specific work is charged on an 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;

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

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
* Balsamic GUI Mockup Tool

# Communications and Engagement

An Agile software development approach will be used for the project. Software will be developed in short iterations, typically 2-4 weeks. 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

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

Removed from the document

# Milestone and Funding Milestone Template

## Funding Estimate

|  |  |
| --- | --- |
| Organisation / Group Name | Anticipated Distribution of EIF Funds (%) |
| Centre for Genetic Epidemiology & Biostatistics at the University of Western Australia | 100% |
|  |  |
|  |  |
|  |  |

## Milestones

**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 | Sub-contract | 31 Jan 2012 | 55 |  |  |  | 20 |
| 2 |  | Established Support Tools & Processes | Support Tools & Processes | 15 Mar 2012 | 8 | 8 |  |  | 1 |
| 3 |  | Project Initiation complete | *Communications plan prepared and sent to NeCTAR (Signed contract + two months).* | 31Mar 2012 |  |  |  |  |  |
| 4 |  | Integrated existing application with AAF Authentication Services | Integrate existing application with AAF Authentication Services | 30 Apr 2012 | 22 | 22 |  |  | 3 |
| 5 |  | Integrated Invoicing & Billing Complete | Integrated Invoicing & Billing Module | 30 Jun 2012 | 15 | 15 |  |  | 40 |
| 6 |  | Initial Production Research Cloud Deployed | Initial Production Research Cloud Deployment | 30 Jun 2012 | 59 | 59 |  |  | 12 |
| 7 | Yes | Funding Milestone 2 |  | 30 Jun 2012 | 55 |  |  |  |  |
| 8 |  | Implemented Data Extraction for Analysis Module | Data Extraction for Analysis Module | 15 Aug 2012 | 6 | 6 |  |  | 2 |
| 9 |  | Implemented Pedigree Storage & Visualisation Module | Pedigree Storage & Visualisation Module | 30 Aug 2012 | 62 | 62 |  |  | 12 |
| 10 | Yes | Funding Milestone 3 |  |  | 55 |  |  |  |  |
| 11 |  | Enhanced Data Linkage & Reporting Module Complete | Enhanced Data Linkage & Reporting Module | 30 Nov 2012 | 48 | 48 |  |  | 9 |
| 12 |  | Implemented Registry Management Module | Registry Management Module | 30 Nov 2012 | 17 | 17 |  |  | 37 |
| 13 |  | Integrated Genotypic Data Management Capability | Integrate Genotypic Data Management Capability | 31 Dec 2012 | 69 | 69 |  |  | 20 |
| 14 | Yes | Funding Milestone 4 |  | 31 Dec 2012 | 55 |  |  |  |  |
| 15 | Yes | Final Admin Closure | *Post-implementation Review (PIR) conducted and sent to NeCTAR.*  *Practical Completion Certificate accepted by NeCTAR.* | 30 Sep 2013 | 31 |  |  |  |  |
| 16 |  | Operations to June 2014 | *Service Levels met and reported to NeCTAR as defined.* |  |  |  |  |  | 189 |

Note that all development software licenses and all development hardware required for the projects will be supplied by The Centre for Genetic Epidemiology and Biostatistics.

# Appendix 1 - Resource Breakdown by Milestone

| **Milestone** | **Associated Deliverable Start Date** | **Milestone Date** | **Assigned Resources** | **FTE** | **Cost** | **EIF Funding** | **Co-investment Funding** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sub-contract signed | Sub-contract | 31-Jan-12 |  |  |  |  |  |
| **Funding Milestone 1** |  | **31-Jan-12** |  |  |  | **$55,000** |  |
| **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 2** |  | **31-Mar-12** |  |  |  | **$55,000** |  |
| **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 3** |  | **30-Jun-12** |  |  |  | **$55,000** |  |
| **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 4** |  | **30-Sep-12** |  |  |  | **$55,000** |  |
| **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 5** |  | **31-Dec-12** |  |  |  | **$55,000** |  |
| **Funding Milestone 6 - Final Admin Closure** |  | **30-Sep-13** |  |  |  | **$31,000** |  |
| System Support | 1-Jan-13 | 30-Jun-14 | Support Analyst (Centre for Genetic Epidemiology & Biostatistics) | 0.5 | $89,500 |  | $89,500 |
|  |  |  | Support Analyst (Centre for MEGA Epidemiology) | 0.5 | $89,500 |  | $89,500 |
| **Milestone Totals** |  |  |  |  |  |  | **$189,000** |
|  |  |  |  |  |  |  |  |
| EIF Salary funds requested |  |  |  |  |  | $306,029 |  |
| Salary co-investment |  |  |  |  |  |  | $325,528 |
| Travel Budget (UWA Centre for Genetic Epidemiology & Biostatistics) |  |  |  |  |  |  | $10,000 |
| Other Expenses (UWA Research Matching Funds) |  |  |  |  |  |  | $10,000 |
|  |  |  |  |  |  |  |  |
| **Total EIF Request** |  |  |  |  |  | **$306,029** |  |
| **Total co-investment** |  |  |  |  |  |  | **$345,528** |
|  |  |  |  |  |  |  |  |
| **Total Project Budget** |  |  |  |  |  | **$631,557** | |

# Appendix 2 - Letters of Support



HOSPITALS DIAGNOSTICS OUTREACH

**ST JOHN OF GOD PATHOLOGY**

PO Box 646

Wembley WA 6913

Tel: 1300 367 674

Fax: 08 9389 7836

www.sjog.org.au/pathology

31 October, 2011

Re The ARK NeCTAR proposal

Dear Paul,

I am writing to formally acknowledge our intent to provide 0.5 FTE of Mr Andrew Mews’ and 0.1FTE of my time to this project. It is essential that we have appropriate database infrastructure and support for our molecular research into cancer as there is no way to manage the genotypic and phenotypic data without it.

Please feel free to contact me if you should need any further information.

Yours Sincerely



Nik Zeps

Dr Nikolajs Zeps

Research Group Leader

[Nik.zeps@sjog.org.au](mailto:Nik.zeps@sjog.org.au)

+61400223097



HOSPITALS DIAGNOSTICS OUTREACH

**ST JOHN OF GOD PATHOLOGY**

PO Box 646

Wembley WA 6913

Tel: 1300 367 674

Fax: 08 9389 7836

www.sjog.org.au/pathology

31 October, 2011

Re The ARK NeCTAR proposal

Dear Paul,

I am writing to formally acknowledge our intent to provide 0.25 FTE of a project officer toward this project if successful. Their role will be focused on the area of biobanking samples from patients enrolled in cancer clinical trials of the cooperative groups.

Please feel free to contact me if you should need any further information.

Yours Sincerely



Nik Zeps

Dr Nikolajs Zeps

Research Group Leader

[Nik.zeps@sjog.org.au](mailto:Nik.zeps@sjog.org.au)

+61400223097



**To**: Mr Paul White

Centre for Genetic Epidemiology and Biostatistics, University of Western Australia

**From**: Professor William Ardrey

WA Institute of Skin Cancer Medicine & Rural Health, and Australian National University

**Date**: October 28, 2011

**RE**: Letter of Support, Cloud Based Informatics Tool

Please accept this letter of support for your grant applications to advance your Java-based, Cloud-based approach to support biomedical researchers.

By way of background, the WA Institute won Royalties for Regions and other grant support to establish a Registry for sufferers of non melanoma skin cancer. The Institute has published its first studies on ocular health, and skin cancer wellness, in rural areas and on remote work sites. We have been working to establish an enterprise based system to capture, catalogue, manage, safeguard and deliver confidential patient data in order to advance this study, and there are no currently ideal tools for this. Through your Centre, your team has commenced collaboration and offered helpful advice, however the system which your Grant funding proposes to support is exactly what GP-led primary care researchers need.

With such a tool available, we would be able to further advance our mission of rural based, GP-led primary care research on other medical issues such as rural women’s health, rural mens wellness, and improved programs for skin cancer prevention, treatment and management ideal for regional areas such as Geraldton. It also is a good tool for commercial medical practices and Australian biotechnology companies, as it adds features and functions to make this sort of primary care research more ‘manageable’ for busy commercial organizations.

Thank you again.

**PROF WILLIAM ARDREY** [**www.wainstitute.com.au**](http://www.wainstitute.com.au) **DIRECTOR, WA INSTITUTE OF SKIN CANCER MEDICINE AND RURAL HEALTH**

**5 HERMITAGE STREET  
GERALDTON, WA 6035**

**Email:** [**director@wainstitute.com.au**](mailto:director@wainstitute.com.au)

**Letter of Support**

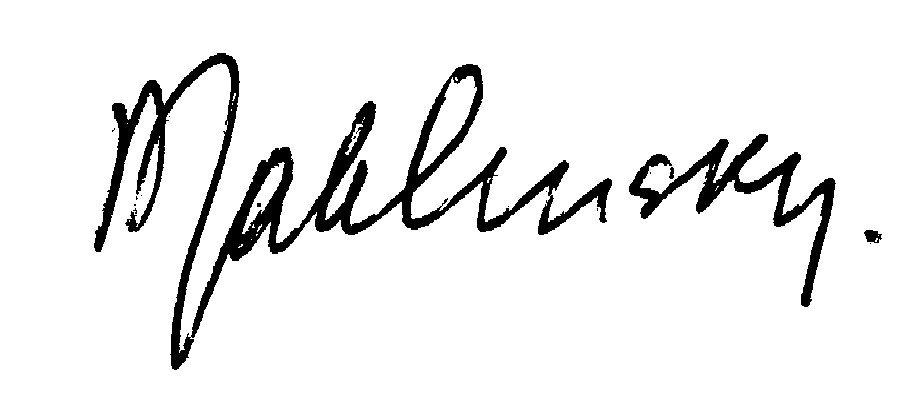
**for The Ark, a “Cloud-based Bioinformatics Tools” proposal of the**

**UWA Centre for Genetic Epidemiology and Biostatistics**

For many years, our teams at the Centre for Clinical Research in Neuropsychiatry and the Western Australian Institute for Medical Research have been involved in studies of the phenotypic and genetic heterogeneity of schizophrenia. Funded by multiple NHMRC grants, the Western Australian Family Study of Schizophrenia (WAFSS) has developed a unique database with extensive information on over 1,000 individuals covering clinical manifestations (including hospital records and video-recorded interviews), cognitive ability (performance on tests assessing different domains of cognition, such as memory, attention, executive function and general intelligence), electrophysiological and neuroimaging data. This is supplemented by a rich collection of biological materials, including DNA, RNA, protein, serum, WBC pellets and cultured transformed lymphocytes. The scope and amount of genetic information is already very large and growing: we have accumulated genotyping data from candidate gene and GWA studies and an increasing amount of sequencing data. Over the years, we have had a very productive collaboration with the team at the UWA Centre for Genetic Epidemiology and Biostatistics, without whose highly professional help and support it would have been impossible for us to manage and use this wealth of information.

The Ark is a major step towards building a tool that will meet future needs for increasingly complicated integration and interrogation of our data coming from multiple diverse fields. Our requirements include expanding capacity (e.g. accommodating additional information on physical morbidity), linking records (e.g. on family members), highly flexible outputs that will save time in generating datasets and reports and allow diverse approaches to data analysis, and of course high security.

The WAFSS collection of data and biological samples is globally unique and The Ark will allow its full use in terms of data mining, current management, and design of future studies.



Winthrop Professor Assen Jablensky 27/10/2011

Director, UWA Centre for Clinical Research in Neuropsychiatry (CCRN)

**Eye**co Pty Ltd

**acn 109 283 892**

31/Oct/2011

Registered Office:

Level 33

William Street

Melbourne, Victoria 3000

Australia

Dear Paul White

University of Western Australia

Please accept this letter of support for your cloud computing initiative.

EyeCo Pty Ltd is aPartner Investigator to the ARC Funded Centre Of Excellence in Vision Science at ANU. We are an ophthalmic drug development company with our first commercial product on the market (for age related macular degeneration). We have developed additional proprietary products which will require clinical trials, data collection, and interface with investigators, regulators, grant supporters and other stakeholders in a confidential way, which also permits data analysis.  The system you are proposing is ideal for the sort of clinical investigations we need to launch in the retinal disease market, and we are very supportive.

Please tell us how we can get involved.  Our team has close linkages with leading universities in Australia and also contributes significantly to peer reviewed research in top journals.

Yours Faithfully

A/Professor Dr Philip L Penfold

Chief Scientist EyeCo Pty Ltd