

NeCTAR Program

Project Closure Document (PCD)

Organisation Centre for Genetic Origins of Health and Disease (formerly Centre for Genetic Epidemiology and Biostatistics)

Project Title **RT029 - Cloud-Based Bioinformatics Tools**

Project Number **RT029**

Stream

Project Sponsor Winthrop Professor Eric Moses

Position Director

Phone 08 6488 6729

Email Eric.moses@uwa.edu.au

Project Manager Travis Endersby

Position Systems Development Manager

Phone 0404 55 63 12

Email travis.endersby@uwa.edu.au

PCD Date 24/03/2014 (first submission date)

PCD Status First submission

FINAL SIGNOFF (when PCD has been approved by all parties)

NeCTAR

Glenn Moloney

Signature

Date

UQ

Signature

Date

Table of Contents

①	PROJECT DETAILS	4
1.1	Project Summary.....	4
1.2	Project Objectives	4
②	COLLABORATION.....	6
2.1	Participating Organisations	6
2.2	Reflection on Collaboration	6
③	USE	8
3.1	Measures and Values	8
3.2	Trends.....	8
3.3	Case Studies and Feedback	8
④	IMPACT.....	10
4.1	Impacts and Benefits.....	10
⑤	SUSTAINABILITY.....	11
5.1	Future Plans	11
5.2	Project Risks and Issues.....	11
⑥	LESSONS LEARNED	12
6.1	Lessons	12
⑦	FINANCIAL	13
7.1	Statement of Income and Expenditure	13
7.2	Asset Register	13
	ANNEXURE 1	14
	ANNEXURE 2	15
	ANNEXURE 3	17

Note to Author

Yellow fields highlight where the Project is required to provide information.

NeCTAR has tried to complete information to the best of its knowledge to assist you.

Please be aware that all responses here will be kept confidential as appropriate.

① PROJECT DETAILS

1.1 Project Summary

1.1.1 Provide a high level summary the project and what it means to Australian researchers.

The project enabled medical researchers to store, and most importantly relate, various forms of medical research and clinical data they produce. The project also aimed to provide a flexible structure in which users with no database experience could design their own custom fields related subjects (people/patients), bio-specimens, bio-collections and longitudinal data sets.

This project was designed with heavy collaboration with a range of scientists (mostly in a range of medical science backgrounds). This lead to a flow that works within the processes used in research labs. Examples of this include the use of uploaders to interact with existing systems as well as a web UI interface aimed to fit in with the laboratory processes.

The projects enables storage, access, reporting and data export of the following types of data

- subject demographic data
- subject custom fields
- biocollection data
- biocollection custom data
- biocollection data
- biocollection data

and 2D visualization of pedigree information. Along with pedigree and genotypic data management.

The objective of The Ark project is to provide a suite of secure, integrated web- based tools that incorporate the majority of the application functionality required to conduct a complex study or clinical trial. The infrastructure aimed to improve research by enabling non-technical researchers the capability to store completely customizable information needed for a research project, around the usual/standard medical research fields without need to understand SQL, programming languages, computer programming concepts as it is all done within a web user interface.

This project grew upon the existing work of the WAGER system (which some call The Ark). Users can now be migrated to The Ark. There is some residual relationship between the WAGER and new Ark systems, but several improvements have been made based on user (and potential user) input.

1.2 Project Objectives

1.2.1 What were the Project's objectives?

The project aimed to build upon existing Ark functionality in order to enable the following;

- Integration with the AAF authentication services;
- An integrated invoicing and billing module;
- A data extraction for analysis module;
- A pedigree (family) data management and visualization module.
- Registry Management functionality for managing participant registries, such as the Australian Twin Registry; and
- A genotypic data management module.

1.2.2 Did the Project objectives vary after the submitted Request for Proposal?

Significant changes to the timeline were necessary following some staffing difficulties and (clearly with retrospect) optimistic timelines. These were all submitted and approved by NeCTAR. One of these changes involved the use of the Genomics Virtual Laboratory which, itself, was behind schedule at the time and based upon that, the decision was made to store meta information on the genetic data (and meta data of the resultant transformations) in The Ark's relational database while the actual genetic information and resultant transformations (together being 200+ GB per sample) will remain stored in a file system.

1.2.3 How did the Project perform against the planned objectives?

The project was able to proceed ahead as expected but there were significant delays in the timeline. Several of the extensions we had hoped to gain grants for were unsuccessful, however this work will hope to be picked up as part of other successful grants utilizing this software. The code/software resulting from the project played a significant part in receiving a significant multidisciplinary Cancer Council research grant. A further collaboration with the University of Melbourne group will ensure further ongoing work also, following the success of their lifepool instance of the project and the uptake of the Project from a collaboration Uni Melbourne have in Korea.

② COLLABORATION

2.1 Participating Organisations

2.1.1 Please add the details of your Project's collaborative partners. Add rows as required.

Organisation / Group Name	Role
Centre for Genetic Origins of Health and Disease (GOHaD)	Project Management. Software development and architecture. Use case development. Software configuration.
The Centre for Mega Epidemiology at Melbourne University	Software Development and use case development. Coordination of Melbourne studies including Lifepool study
St John of God Healthcare	Input regarding business use cases. Facilitated User Acceptance Testings.
WA DNA Bank	Active users, beta testing and User Acceptance Testing

2.2 Reflection on Collaboration

2.2.1 Did the collaboration exist before the Project came into being? If yes, did the Project strengthen the collaboration?

This collaboration was an outstanding success. The individual skills available in both teams complemented each other extremely well. Additionally having teams on both sides of the country gave an opportunity for more personal contact with potential users. There will be future collaboration in the future between GOHaD and MEGA.

The MEGA team played a significant part in relaying the usage of the software from the lifepool project's perspective. The developers on both sides of the country were able to bounce ideas off each other and provide training in areas needed.

This collaboration had not existed prior to the beginning of this project. However, as an update, this collaboration has continued after the project and has spawned an expansion of this project plus a new relationship with the MEGA group on a successful Cancer Council Grant. Further to this, the MEGA team is utilizing the UWA team on work for Seoul National University (SNU) project which will utilize The Ark open source software to maintain/access Korean studies.

In regards to the Obiba relationship, there had been previous information exchange (with our previous PM and current developer) in either direction as they had similar software infrastructures and projects. This also involved a trip to Canada, after which, they proposed (and the team accepted) certain elements (patterns) of the software engineering structure (including the use of Apache Wicket as a web framework). There was not a significant amount of collaboration since my arrival on the project approximately 2 years ago.

In relation to the COSA collaboration; The most significant part of that relationship during my presence was the outstanding advice given by COSA Board member and Steering Committee member, Professor David Goldstein throughout the project on how the project could have the greatest clinical relevance. I have not been able to establish that there was any previous collaboration that existed before the Project came into being.

2.2.2 How did collaboration strengthen the Project?

The MEGA team played a significant part in relaying the usage of the software from the lifepool project's perspective. The developers on both sides of the country were able to bounce ideas off each other and provide training in areas needed.

2.2.3 Did the Project's collaborative partners deliver what was promised?

All partners offered significant input to the project. St John of God Healthcare are yet to provide the cash contribution as originally promised. GOHaD/UWA are covering this cost of staffing for this support period while we try to mediate some kind of solution with SJOG.

2.2.4 Did the number of collaborative partners increase throughout the project?

No. Although several groups will have additional collaborative projects resulting from this project.

③ USE

3.1 Measures and Values

3.1.1 How did the Project measure success?

The project had action items mandated at the quarterly steering committee meetings. These were always met. The completion of modules and UAT of each module was also how the project's components were shown as successful. The overall success of the project could be measured by the number of users, and the extent to which the users were able to use the capabilities of the project. These numbers steadily increased with time as shown below

3.1.2 Project measures – detail.

All supporting graphs, data etc in **Annexure 1**. The original document merging the 9 reports and all tables and graphs are available in the submitted Excel document attached – if you require a hard copy of all graphs, please ask. So there are now 4 additional tables and further additional graphs which illustrate the increased uptake. (In addition to the “Measures” tabs from the 9 original reports – all in the one excel document).

3.2 Trends

3.2.1 How did uptake improve over time?

All supporting graphs, data etc are available in **Annexure 1**.

There was a general uptrend, which has continued to this date.

3.2.2 What are the future projections of this uptake?

I believe there will be a rapid increase in uptake of this project going forward. When I took the project over the steering committee agreed it was time to catch up on deadlines and hold off on actively chasing new users for some time. We have several users throughout Europe, Asia, Africa and Australia running their own instances of this open source software and several that have intentions of maintaining managed instances going forward. There will be much more active “marketing” via demonstrations and potential-user education of the software going forward.

3.3 Case Studies and Feedback

3.3.1 Success Stories

See ann. 2.

3.3.2 What was the general feedback from the User community toward your Project?

Users have certainly been happy/satisfied relative to the other offerings they have tried. We have received positive feedback from users regarding the user friendly flow of the application (the has been repeatedly refined through out the process).

I would suggest we could improve the way in which users themselves could be educated on what they can and cannot do without support. Support time has been a little bit limited while stretching for objectives. Going forward it will create a better expeience for the users and potential users to have good documentation and video guides on how to use the software.

There is certainly a list of JIRA requests still in existence, so users could be happier, but even our most heavy JIRA users are more satisfied with what we have than their previous offerings.

④ IMPACT

4.1 Impacts and Benefits

4.1.1 Describe examples of the impact and benefits that your Project has had on research.

E.g. Tangible benefits; Citations; Has your work led to further grants or funding? What future benefits do you anticipate?

There have already been numerous new users and collaborations (nationally and internationally) lined up as part of this as mentioned before. The most significant grant has been a multi-disciplinary cancer council grant valued at 1.2 million dollars, which will to use the software from this project as it's means to store all data related to predictive modelling of colorectal cancer risks.

Official cancer council page;

<https://www.cancerwa.asn.au/research/successfulresearchfunding/researchprojectsandinitiatives/>

UWA Press release;

<http://www.news.uwa.edu.au/201405166682/cancer-council-wa/cancer-council-wa-capacity-building-and-collaboration-grant-12-m-awar>

We would happily work together with Nectar on a more coordinated, targeted, NeCTAR/Ark-centric press release based on this and other successes as they are achieved.

This software is also the part of many current grant applications from MEGA/University of Melbourne and GOHaD/UWA and the software is being used by groups throughout the world that are seeking an open source solution to their research data storage needs. There will certainly be a flow on affect from this project. University of Melbourne have also agreed to fund GOHaD developers to further the work of this project and assist in setting up instances for international collaborators.

⑤ SUSTAINABILITY

5.1 Future Plans

5.1.1 What are your plans for the future of the infrastructure?

As part of the grant mentioned in 4.1, there is plans to expand the capabilities of the project to include greater integration with analysis pipelines. Our PhD student at MEGA will also be expanding on this work as part of his PhD. There are several other grants that have included money towards bioinformatics support that could enable further development. We are very keen to create a community of both users of the software and contributors in order to allow the software to continue on well beyond this projects end.

5.1.2 How will the future plans be supported?

The community inclusion element will be supported by user education. The PhD student and funded grant are all in motion and each have their own mechanism to measure success or failure in the goals they are trying to achieve.

5.2 Project Risks and Issues

5.2.1 What are the ongoing Project risks?

Enabling funding to the extent that we can continue to maintain a public face and leadership to get through the initial momentum needed to create community collaboration in the open source software.

⑥ LESSONS LEARNED

6.1 Lessons

6.1.1 What went well?

Collaboration between development teams. Coordination with a complex group of users with different requirements

6.1.2 What did not go well?

Timelines were not managed well. My personal opinion is the original timeline was too tight in the first place. This we worsened by a cut to the proposed budget and then further by delays in staffing at times. Further to this those involved in the project certainly needed to say “no” more often to feature requests even if they made for a better project, because new features always come at a time cost.

6.1.3 What would you do differently?

I believe certain interim goals would need to be laid out. Particularly when dealing with development teams that do not have a lot of experience in laying out estimations. Even experienced developers tend to not fully account for unforeseen circumstances, but this is a very common trend. Smaller targets along the way, set out by the developers could make their initial estimate more accurate in the first place – and help give clues when they are falling behind.

6.1.4 If not already outlined above, how do you feel the NeCTAR project supported you and your Project through the initial call for proposals, the negotiation, the development, the communication/marketing, and this final report? How could it improve?

The NeCTAR team have been responsive. Having come into the role of project manager a significant way through this project, I probably was not as prepared as I could have been. But having a more definitive layout of what exactly was required of me along the way would have helped. That being said, communications were outstanding to help me catch up on what I had missed. Nigel Ward was also very useful in his suggestions as a steering committee member.

⑦ FINANCIAL

7.1 Statement of Income and Expenditure

7.1.1 Please provide an Audited Detailed Statement of Income and Expenditure in **Annexure 3**.

It must include:

- A definitive statement as to whether the financial accounts are true and fair, and a statement of the balance of your account
- An audit statement that the Funding was expended for the Project and in accordance with the Contract.

The Statement of Income and Expenditure must be audited and certified by:

- a) Your Chief Financial Officer and
- b) The Senior Executive Officer who has primary responsibility for managing the audit functions in your organisation.

7.2 Asset Register

7.2.1 Please attach a certified copy of the Asset Register.

You can use the Asset template included in your Nectar Reports and include a statement similar to the one provided below.

I hereby certify that:

- a) All assets were acquired and created as part of the Nectar Project and in accordance with the Terms and Conditions between Nectar and UWA RT029.
- b) Assets above are described in compliance with Section 4.2 of the Terms and Conditions – Specific Provisions regarding Location, Ownership and Value.

The Asset Register must be certified by the Chair of the Steering Committee.

Project Sponsor

Signature

Date

ANNEXURE 1

Measures and Values over time

This graph has now been replaced with an highly detailed excel spreadsheet combining all reports submitted to nectar and graphs representing those 9 reports' data. See attached document.

ANNEXURE 2

Case Studies

Case study 1 – The lifepool project;

- NBCF funded to recruit 100,000 Victorian women for breast cancer research
- The Ark: open source web-based research data management
- Currently:
 - NSP hosting an Ark v1.1.1 instance
 - ~45,000 participants' data is managed
 - Demographic / consent / scanned documents / biospecimen
 - ~50 research users
 - Relied upon daily, from multiple locations
 - *“The Ark, provides an elegant solution to the ethical issues surrounding electronic storage of participant data. The system is a crucial part of the project’s day-to-day operations” Lisa Devereux, Project Manager, lifepool - Peter MacCallum Cancer Centre*

Case Study 2 – The WA DNA Bank (WADB)

The Western Australian DNA Bank (WADB) is a national processing and long-term secure storage facility for biospecimens (including DNA, RNA, serum and plasma) that have been collected from donors who have participated voluntarily in one or more medical research projects. The WADB itself does not recruit these donors, but processes and stores biospecimens for medical researchers who have collected a sample from consenting donors.

The WADB is a world-class resource helping Australian medical researchers to more easily conduct large-scale health studies and so speed up their search for life-changing discoveries.

The project, initially funded by the Federal Government through the National Health and Medical Research Council of Australia (NHMRC) and currently by the University of Western Australia and NHMRC, provides scientists with a state-of-the-art facility to process and store DNA, RNA, serum and plasma samples needed to undertake critical medical research into common diseases such as cancer, schizophrenia and heart disease.


The WADB brings together large numbers of DNA samples under one roof in a highly secure way, and because of this, it is of extraordinary value to the nation's scientific community. By helping to aid Australian researchers' access to larger collections of DNA samples, which can often be expensive and time consuming to collect, the outcomes of their studies are likely to be more reliable, therefore having an increased impact on the health of our community.

It is also be an important tool for current and future large population-based studies in WA, such as the existing Western Australian Family Study of Schizophrenia and the Inherited Retinal Disease Study. [Click here](#) for a list of collections currently maintained by the WA DNA Bank.


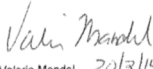
The initiative, which builds upon pre-existing facilities, is based at two separate sites in Perth, Western Australia, which maximises the safety of the DNA samples.

WA DNA Bank hosts data in The Ark and going forward to do so for all the studies under its control.

ANNEXURE 3 Audited Detailed Statement of Income

 THE UNIVERSITY OF WESTERN AUSTRALIA		Registrar's Office	
		Research Grants Office Research Services M459 The University of Western Australia 35 Stirling Highway Crawley WA 6009 Phone +61 8 6488 7644 Fax +61 8 6488 1075 Email leah.simeon@uwa.edu.au http://www.research.uwa.edu.au CRICOS Provider Code: 00126G	
University of Melbourne ex National eResearch Collaboration Tools & Resources NeCTAR			
Cloud Based Bioinformatics Tools Dr Paul White RA/1/1781/2 01205 / 49000100			
STATEMENT OF INCOME AND EXPENDITURE for the period 31/07/2013 to 31/12/2013			
	Cash	Inkind	Total Cash + Inkind
Opening Balance as at 31/7/2013			-113,612.60
Income			
In-kind contributions		107,055.00	
Income Received from NRP	104,000.00		104,000.00
Total Income			-9,612.60
Expenditure			
Salaries	28,422.41	107,055.00	
Computing Costs	0.00		
Lab, Workshop & Medical	0.00		
Professional Fees & Consulting	0.00		
Staff Training	0.00		
Travel	0.00		
Other Expenditure	60.78		
Minor Equipment Purchase	0.00		
Total Expenditure			28,483.19
Balance for period ending 31/12/2013			-38,095.79

I certify that the above Statement represents an extract from the accounts and records of The University of Western Australia, for the purpose of preparing the Statement for the activity for the period 31/07/2013 to 31/12/2013

Prepared by:  19/3/14 Leah Simeon Research Finance Officer Research Grants Office	Reviewed by:  20/3/14 Valerie Mandel Finance Manager, Research Finance Research Grants Office
---	---

ANNEXURE 4 Asset Register

Cloud Based Bioinformatics Tools
01-Oct-13
31-Dec-13

Now in the table below for any assets delivered
for further information.

Asset Type	Name	Version Number	Owner/ Custodian	Value	Location	Local Asset Tag	Status
Software	Integrated Billing & Invoicing	1.1.1	UWA	\$54,768.00	Atlassian Subversion Repository	ARK_1.1.1	Production
Software	Questionnaire & LIMS Modules	1.1.1	UWA	\$42,374.00	Atlassian Subversion Repository	ARK_1.1.1a	Production
Software	Registry Module	1.1.1c	UWA	\$42,374.00	Atlassian Subversion Repository	ARK_1.1.1b	Production
Software	Initial Production Research Cloud Deployed	1.1.1c	UWA	\$97,714.00	Atlassian Subversion Repository	ARK_1.1.1c	Production
Software	Initial Production Research Cloud Deployed	1.1.1d	UWA	\$86,780.00	Atlassian Subversion Repository	ARK_1.1.1d	Production
Software	Initial Production Research Cloud Deployed	1.1.1d	UWA	\$86,781.00	Atlassian Subversion Repository	ARK_1.1.1d	Production
Software	Improvements on same tag	1.1.1e	UWA	\$34,303.00	Atlassian Subversion Repository	ARK_1.1.1e	Production
Software	genotypic module internal release	1.1.1e	UWA	\$38,089.00	Atlassian Subversion Repository	ARK_1.1.1e	Production

All assets listed above are available at our open-source repository location (which has since moved from Atlassian to google-code hosting). The url is <https://code.google.com/p/ark-informatics/>

Given space limitations please asset that UWA in the above graph refers to the certification below which states;

I hereby certify that;

- All assets were acquired and created as part of the Nectar Project and in accordance with the Terms and Conditions between Nectar and The University of Western Australia RT029 Cloud-based Bioinformatics Tools
- Assets above are described in compliance with Section 4.2 of the Terms and Conditions – Specific Provisions regarding Location, Ownership and Value.

Travis Endersby, Project Manager

Eric Moses, Project Sponsor