



# **ARCS Annual Business Plan 2008-09**

**1 July 2008 to 30 June 2009**

**Professor Anthony G Williams  
Executive Director of ARCS**

**<http://www.arcs.org.au>**

## Contents

1. Introduction and background.....	3
2. Current Status of ARCS .....	3
2.1. Structure and Governance.....	3
2.2. Management and Processes .....	4
2.3. Progress since the beginning of the NCRIS Funding Period .....	5
2.3.1. Collaboration Services .....	5
2.3.2. Operational Services: Systems services.....	6
2.3.3. Operational Services: Data Services .....	6
2.3.4. Projects .....	7
3. Planned Activities for the Period.....	9
3.1. Goals for the Period: Executive Summary .....	9
3.2. Services.....	10
3.2.1. Collaboration Services .....	10
3.2.2. Operational Services: Systems services.....	10
3.2.3. Operational Services: Data Services .....	11
3.2.4. Projects .....	11
3.2.5. Authorisation Services.....	16
3.3. Milestones .....	17
3.4. Measures.....	18
3.5. Communication and Outreach .....	18
4. Budget for the Period .....	19
Appendix A: ARCS Budget for the Period 2008-9 .....	21
Expected Income for the Period .....	21
Current Plan for Expenditure .....	22
Appendix B: ARCS Position Description Template.....	24
Appendix C: ARCS Delegation of Powers.....	26
Appendix D: Resolutions re Potential Conflicts of Interest.....	27
Appendix E: NeAT Project Plan Summaries.....	28

## 1. Introduction and background

This document is the Australian Research Collaboration Service (**ARCS**) Annual Business Plan for the period 1 July 2008 to 30 June 2009. ARCS is an unincorporated collaborative venture formally constituted on 15 November 2007 for the purposes of providing long-term eResearch support to the Australian research community. ARCS activities and governance processes are defined in the **ARCS Collaboration Agreement**. ARCS activities include, but are not limited to, the provision of national eResearch support services in the form of: interoperability and collaboration infrastructure and services; authorization services; services to the National Collaborative Research Infrastructure Strategy (**NCRIS**) capability areas; services to other national discipline-based research communities; and other related activities as appropriate.

The ARCS members, referred to hereafter as the Members of ARCS (**MARCS**), are: the Australian Centre for Advanced Computing and Communications (**ac3**) in New South Wales; the Australian National University (**ANU**) in the Australian Capital Territory; the Commonwealth Scientific and Industrial Research Organisation (**CSIRO**) nationally; eResearch SA (**eRSA**) in South Australia; **IVEC** in Western Australia; the Tasmanian Partnership for Advanced Computing (**TPAC**) in Tasmania; and the Victorian Partnership for Advanced Computing (**VPAC**) in Victoria. The Queensland Cyber Infrastructure Foundation (**QCIF**) is expected to sign the ARCS Collaboration Agreement and become one of the MARCS in the near future and **INTERSECT** is being formed in NSW and will soon join as the MARCS replacing and subsuming ac3.

The NCRIS funding period is from 1 July 2007 to 30 June 2011 and the primary initial purpose of ARCS is to deliver the Interoperability and Collaboration Infrastructure (**ICI**) component of the NCRIS capability area known as Platforms for Collaboration (**PfC**) to the Department of Science Education and Technology (**DEST**). The ICI funding is \$20 million over four years with \$14 million allocated to ICI-related operations and \$6 million for the ARCS components of projects selected and approved by the National eResearch Architecture Taskforce (**NeAT**).

Subsequent to the formation of ARCS, the role of DEST relevant to ARCS has been taken up by the Department of Innovation, Industry, Science and Research (**DIISR**). DIISR has recently indicated that additional funding of \$2 million will be provided to ARCS to deliver Authorisation Services and a further \$500k will be provided to assist in the delivery of Collaboration Services. The delivery of ICI services by ARCS to the Commonwealth is specified through an **ICI Funding Agreement**, which is included as a Schedule in the ARCS Collaboration Agreement. The additional \$500k for Collaboration Services will be delivered by an amendment to the ICI Funding Agreement. A separate **Authorisation Funding Agreement** will be put in place between DIISR and ARCS and will be included in the ARCS Collaboration Agreement as an additional Schedule to cover the delivery of the Authorisation Services.

## 2. Current Status of ARCS

### 2.1. *Structure and Governance*

ARCS is governed by the processes set out in the ARCS Collaboration Agreement. All ARCS activities are subject to the governance of the Executive Committee, which includes a nominated representative of each of the MARCS, the Executive Director (ex-officio), a member appointed by Affiliates and up to three Independent Members. No Affiliates have yet

joined and no Independent Members have yet been appointed although this is anticipated to change during the coming business year. There is an option for an Independent Member to be appointed and remunerated as Chair and this is currently being actively considered by the Executive Committee. The Executive Director is appointed by and reports to the Executive Committee.

As ARCS is an Unincorporated Collaborative Venture it must operate through a legal agent. ARCS refers to this role as the **Lead Agent** and this is currently VPAC although this can change if necessary. The ARCS Collaboration Agreement states that **the Lead Agent will act, all times, under the direction of the Executive Committee in relation to its activities as Lead Agent.** It is also possible in principle to have different MARCS act as legal agents for each individual contract that ARCS has with an external entity as well as a separate MARCS acting as Lead Agent for the ARCS collaboration itself. At present VPAC is the Lead Agent for the collaboration as well as being the legal agent for the ICI Funding Agreement with the Commonwealth. A different legal agent could be used for the Authorisation Funding Agreement with the Commonwealth, if desired.

As outlined in the Appendix entitled 'Resolutions of Potential Conflicts of Interest': ARCS activities are by definition subject to the governance procedures of ARCS and as such will involve the Executive Director and the Executive Committee in their development and their delivery; ARCS Business by definition has associated monies flowing through the ARCS accounts; and the MARCS have agreed that **ARCS is the preferred vehicle for the delivery of national e-Research support services** and to work together to pursue opportunities for sustainability and growth of ARCS.

The structure of ARCS was adjusted to its new form during Q1 of 2008. It consists of Collaboration Services, Operational Services, Authorisation Services and Projects. Operational Services in turn contains the Systems services Team and the Data Services Team with each headed up by a Team Leader. The Authorisation Services activities will not begin formally until the new business year begins on 1 July 2008.

## **2.2. Management and Processes**

All substantial new business decisions and directional strategies are decided through the regular processes of the Executive Committee in a manner consistent with the ARCS Collaboration Agreement. For ICI-related business the activities of ARCS are in addition bound by the conditions of the ICI Funding Agreement. A similar situation will exist for Authorisation Services once the Authorisation Services Agreement is in place. The Executive Director is responsible for the day-to-day management of ARCS with the assistance of the various Managers and Team Leaders. The Delegation of Powers within ARCS has been agreed and is attached as an Appendix. The Executive Director is the Line Manager for each of the Managers and each of the Managers are in turn the Line Managers for the ARCS-funded staff and the activities in their respective area. Since some staff may work across more than one area, they may have more than one Line Manager.

The Management Group is convened by the Executive Director and includes the Operational Services Manager, the Projects Manager, the Collaboration Services Manager and the Authorisation Services Manager. The Collaboration Services Manager and the Authorisation Services Manager positions are soon to be advertised and it is anticipated that they will be filled during July 2008. The Manager positions are professional positions with significant responsibilities to ARCS and appropriate recruitment processes will be used to fill and, as needed, replace these positions. Processes similar to those used to recruit the Executive Director will be used. The remuneration for the Manager positions will be determined by ARCS, based on skills and experience and on a case-by-case basis. The Local Host for a

Manager can be any of the MARCS or any of their partner institutions. The Local Host offering to host the Manager will second him/her to the local MARCS if they are not already an employee and then on-second them to ARCS itself. The local MARCS will receive the standard ARCS funding for a Manager in return for providing the hosting for the Manager. ARCS will ensure that the salary plus on-costs at the Local Host does not exceed the standard ARCS funding for a Manager. ARCS will work with the host MARCS and the Local Host institution to ensure a consistent relationship between ARCS Management processes and needs and the Local Host human resources requirements.

Coordination across Services, Teams and activities is ensured through weekly EVO videoconferences of the Management Group and Team Leaders. Each Team also has weekly or fortnightly videoconferences. In addition, the Position Description Template included as an Appendix makes it clear that ARCS activities carried out by ARCS-funded staff are under the effective control of the Executive Committee governance through the ARCS Line Management arrangements. This is necessary in order to ensure that all ARCS Activities are subject to the governance procedures of ARCS. ARCS funded staff may also deliver non-ARCS incidental services to the Local Host under the direction of the Local Host line manager

### ***2.3. Progress since the beginning of the NCRIS Funding Period***

#### **2.3.1. Collaboration Services**

The Collaboration Services group was started in Q2 2008. Prior to the formation of this group, there was a User Services team which operated under Operational Services. The User Services group has worked closely with the Operational Services Group to further the development and deployment of authorisation and job submission toolkits, Grix and Gris. Additionally during the period many improvements have been made in the quality of ARCS' Information Services, based on MDS. It is now possible to make use of the Information Service to determine the necessary information needed to submit grid jobs. Use of this Information Service has been developed into ARCS' job submission tools. Further work has also been put into ARCS' Testing Services, based on INCA, to provide a robust framework for testing the availability and correctness of a wide range of ARCS' production services. Responsibility for the above technologies has been merged into Systems services with the formation of this and the Collaboration Services groups.

The formation of the Collaboration Services group has seen a number of new web and video based tools being deployed within ARCS for the use of the Australian research community. Web based collaboration tools, such as Sakai, have been deployed and sites set-up for a number of Australian research groups and for DIISR to support the NCRIS Roadmap Review. Feedback from these groups has been very positive to date, with one group of medical researchers using these tools to collaborate with researchers in Japan. A Plone site has been set up for ANDS and for the International Lattice Data Grid (ILDG) Project. Progress has also been made in the areas of video collaboration, with work under way on an agreement with Caltech to provide EVO support in Australia. EVO is being extensively used by ARCS itself as well as ANDS and a number of research groups such as AuScope. In order to make the Access Grid more robust, Quality Assurance processes have been initiated including certification for AG sites and we have worked with AARNet to improve the performance of multicast. In addition an email based Help Desk has been established, with work well under way to have a 1-800 telephone Help Desk operational in the beginning of the 2008-9 year.

### **2.3.2. Operational Services: Systems Services**

The Systems Services Group incorporates the old APACGrid Compute Infrastructure Project and a number of activities that were in the hands of the APACGrid Portals Project. During the ARCS Establishment Period, the User Services and User Support groups came into existence and have now been merged into either Systems Services or the newly formed Collaboration Services group. Systems Services, while known previously under other names, has been responsible for operating and further developing the Compute Grid infrastructure that is now installed at fourteen HPC sites around Australia and New Zealand. Until recently this group's activities included providing support for ARCS' Collaboration Services activities and currently continues to provide the Authentication and Authorisation activities used by Grid and related researchers in Australia.

Daniel Cox was appointed Team Leader in recognition of the leadership he showed to the APACGrid Grid Admin group. Initially the Systems Services concentrated on transitioning from an APACGrid model to the ARCS model and from range of developmental models to production quality services. To this end the re-badgeing processes are effectively completed and web, wiki and Trac content ported and reviewed. The Grid and other system resources are now increasingly robust, with a new focus on Quality Assurance and appropriate change management processes. A help system is now in place that is already widely used.

The Compute Grid has seen well over a hundred and 175,000 CPU hours delivered to researchers in the second half 2007 and already the first quarter of 2008 saw over 160,000 CPU hours. During that time we saw the completion of the transition from a remarkably reliable but still officially developmental Grid to a production quality service placing emphasis on customer satisfaction and usability. The Certificate Authority has 620 current certificates issued and the VO Management System has 333 researchers subscribed. The certificate management tool, Grix, was developed early in this period and is now widely used and considered stable. The adoption of this tool by researchers is a good example of ARCS' approach of making things easier for end users by hiding the complexity but still providing the power users the capability of leveraging their skills. Considerable progress has been made with the recently developed Grisu tool kit that allows easy creation of application specific desktop graphical user interfaces for remote job submission using ARCS grid infrastructure. Already much of the existing grid usage is now being launched via Grisu.

### **2.3.3. Operational Services: Data Services**

The ARCS Data Team arose out of the APACGrid Data project, referred to as the Information Infrastructure Project at the time, and was formally set up as a team under ARCS Operations in July 2007. Stephen McMahon from the ANU, who had been active in the APACGrid, was appointed as Team Leader. His role is to provide technical leadership in the range of activities that ARCS has determined the Data Team should conduct. While membership of the team has varied over time, it currently involves eight people and at least a half time commitment from all states of the Commonwealth.

The Team initially inherited much of its technology from the APACGrid, but in ARCS it has needed to work much more closely with other ARCS groups and has done so very successfully. The Data Team has clearly defined, from a user point of view, the data services that it is appropriate for ARCS to provide and is working on the underlying technology that these services depend on. A range of reliable and regularly tested services now exists between 11 sites and the groundwork is in place to expand both the number of services and the number of involved sites. The group has regular team meetings and is active with other ARCS groups and, importantly, the wider research community. Concentrating on data staging, data movement, data replication and distribution the team has documented some good clear examples of how end users can make good use of our capabilities on the web site <http://www.arcs.org.au>.

### 2.3.4. Projects

Much activity in Proto-NeAT projects and related User Support activities has occurred during the first year of ARCS ICI funding. This is briefly outlined below:

**Virtual Observatory Web Services for Astronomy Data (ANU and ac3):** The astronomy community has developed international standards for metadata, data formats and web interfaces for searching astronomy data through the International Virtual Observatory Association (IVOA). The ANUSF hosts the Massive Compact Halo Objects (MACHO) data set, an important astronomical data resource consisting of 127,000 images (~7 TBytes) collected from 1991 to 2003 by the ANU Mount Stromlo 50" telescope, and an associated database of 75 million stars. This project has made the MACHO data set available using IVOA standards, which required developing software to process the images and database to generate standard IVOA metadata, and developing web services for searching and accessing the data using standard IVOA interfaces. Work has started on developing prototypes of similar services for data from the SkyMapper project, which generates over 1 TByte of data per night. The deployment of IVOA services for the Sydney University Molonglo Sky Survey (SUMSS) data is currently underway. A related project at ac3 has developed IVOA services for two telescopes managed by UNSW.

**Bioinformatics Data Sets Available Over The Grid (QCIF):** Genomic and proteomic data sets including GENBANK, GO, PDB, UNIPROT have been stored at the Queensland Facility for Advanced Bioinformatics (QFAB) using Storage Resource Broker (SRB), and made available via SRB client tools and through standard tools such as the Sequence Retrieval System (SRS) and the UCSC Genome Browser. Mirroring and synchronization of required data sets has been implemented. Authentication and authorization mechanisms using Shibboleth are being deployed, in collaboration with the ARCHER team at JCU.

**Bioinformatics Portal and Workflow Development (SAPAC/eRSA):** A web portal was developed for submission of BLAST sequence alignment jobs for large numbers of sequences, by splitting the job into multiple tasks and using Nimrod to run the jobs concurrently using multiple processors. The portal has been deployed at SAPAC and used by the Australian Centre for Plant Functional Genomics (ACPFG) and other research groups. A variety of workflow solutions were investigated for comparative analysis of genomic data for agriculturally significant grasses such as wheat and barley. A custom workflow solution was developed that uses grid job submission. The Grisu grid job submission portal was customized for use with phylogenetic applications (Mr Bayes, PAUP and BEAST) and used by researchers at University of Adelaide. A web-based customized microarray data analysis portal was developed for ACPFG.

**Grid-enable Sequence Assembly Tools (iVEC):** This project investigated available sequence assembly tools for traditional and next generation genome sequencing instruments (such as 454 and Solexa), selected the open source tools that could be run most effectively on parallel computers, and developed software to enable them to be accessed via standard grid middleware supported by ARCS. A web-based workflow system (YABI) was developed to allow users to easily specify workflows for sequence assembly. The tools were used by the Centre for Comparative Genomics for assembly of 454 sequences for the wheat genome, as part of the International Wheat Genome Sequencing Consortium (IWGSC), and Ixodes Scapularis, for the Beef CRC.

**Proteomics Laboratory System (ac3):** A laboratory information system was developed that enabled researchers to upload information about proteomics experiments to a database using a web interface running on a desktop or laptop computer or a wireless PDA. Information included photographs of the experiment taken using the PDA with associated

metadata. A framework for a digital mentor expert system was also developed and trialed. The system has been deployed and used at UTS.

**Collaboration Services Infrastructure (QCIF):** Access Grid (AG) is widely used for collaboration and videoconferencing by researchers and by the MARCS and other participants in PfC. QCIF runs many of the national AG services and is involved in the international development of AG software and services. This project worked on improving the reliability of AG in Australia and developing additional functionality. Work on improved reliability included the development, documentation and operation of a quality assurance program for AG nodes; monitoring and improvement of multicast network performance over AARNet; improving fault tolerance by reducing reliance on single servers (e.g. the Jabber server used for the AG chat client); improved unicast bridge support and deployment; and bug fixing in AG code releases. Work on improved functionality included adding support for new video codecs and high-definition digital video; and a shared whiteboard.

**gLite Adoption (VPAC):** gLite is the grid middleware used by the Enabling Grids for E-Science (EGEE) project, predominantly in Europe. The main project using gLite middleware is the CERN Large Hadron Collider (LHC) Grid. ARCS does not currently support gLite, but it is required by Australian researchers involved in international experimental high-energy physics collaborations including ATLAS (one of the LHC detectors) and Belle. This project supported the development of expertise in gLite, the deployment of an EGEE node using gLite middleware at the University of Melbourne, and the development of a plan for a broader deployment of gLite by ARCS, which is expected to occur in Q2 2008.

**Grid Enabling Training Courses (ac3):** This project supported the development of training material in grid computing.

**Integrated Environment for Computational Chemistry (ANU, ac3, iVEC):** Computational chemists are probably the largest users of high-performance computers at the MARCS and NCI. This project has developed a desktop application called JMolEditor that allows users to more easily specify and submit computational chemistry jobs to remote supercomputers, monitor the progress of the jobs, and visualize the results (e.g. molecular structures and molecular dynamics trajectories). This phase of the project focussed on developing additional visualization functionality, adding support for a wider range of programs and input and output data formats, providing more complete grid job submission and monitoring capability, improving documentation and tutorials, and talking to many existing or potential users of the software to identify and address their requirements. The software is currently used by several research groups, and is used in undergraduate and honours computational chemistry courses at the University of Adelaide.

**Neutron and X-Ray Data Grid (ac3):** The project is to undertake a pilot implementation of the data management and access system that is being proposed for NCRIS 5.3, which is based on the system being developed by the UK Science and Technology Facility Council (STFC) for its neutron and synchrotron facilities. The pilot project targets the OPAL neutron facility at ANSTO and X-ray diffraction instruments at the University of Sydney. Services being developed include metadata capture at source, generation of data in standard interchange format (imgCIF and NeXus/HDF), storage of data using Storage Resource Broker (SRB), a metadata catalog (ICAT) based on the STFC Scientific Metadata Model, and a web portal service (the STFC DataPortal) providing search and retrieval across facilities and platforms. The project is also investigating related work done in the ARCHER project, the use of Shibboleth for authentication and authorization, and iRODS as a successor to SRB.

**Nimrod Support (VPAC):** Nimrod is a tool used for computational parametric studies, where multiple compute jobs with different input parameters are run concurrently on many

processors. A version of Nimrod has been developed to submit jobs using the Globus Grid Toolkit middleware supported by ARCS. A web portal for Nimrod has been developed to allow users to easily specify, submit and monitor parametric jobs. This project supported the deployment, maintenance and user support of a Nimrod portal at VPAC that can submit jobs to compute resources at multiple MARCS sites; the development of documentation to assist users in making use of the portal; the development and presentation of tutorials on the use of Nimrod; and the development of additional functionality and bug fixes for Nimrod and the Nimrod portal.

### 3. Planned Activities for the Period

#### 3.1. Goals for the Period: Executive Summary

*The continuing ARCS Mission is to provide long-term eResearch support services for the Australian research community with a particular focus on interoperability and collaboration infrastructure, tools, services and support. Careful attention will be paid to the pro-active delivery of customer-oriented, production tools and services with appropriate change management and quality assurance processes in place.* The primary tools and services to be provided by ARCS in 2008-9 include:

- ARCS Help Desk - Email: help@arcs.org.au and Telephone Service (1-800 number becoming available very soon)
- Video Collaboration Tools, Services and support - including Access Grid, EVO, Skype, Jabber and others
- Web-based Collaboration Tools, Services and Support - including Sakai, Drupal, Plone, Google Calendar, Google Apps and others
- ARCS Data Fabric including Data Storage, Access and Movement Tools, Services and Support
- Grid and Remote Submission Tools Services and Support
- Authorisation Services – providing expertise, exemplars and best practice tools for eResearch service providers to link their authorization needs with AAF authentication and other service providers.

Over time ARCS intends to develop and deliver additional tools and services including:

- Remote Instrumentation and Sensor Network Activities
- Consultation Services
- Training Courses and Services.

ARCS is also actively working to provide needed eResearch collaboration tools and services for ANDS and for other NCRIS capability areas such as ABIN.

As mentioned above, a particular focus for the coming Period will be the **formal establishment and uptake of Quality Assurance (QA) and Change Management processes to harden up our Production Services** and improve our ability to provide a robust core services platform for users and developers from the various research communities. **These processes will also be applied to each of the NeAT Projects.**

An effort will also be made over the period to develop a formal relationship with the University IT Directors through CAUDIT as their cooperation and positive engagement is central to the delivery of many of the ARCS Services, including for example Authorisation and Collaboration Services.

ARCS will embrace existing and commodity-like tools and service wherever they are available, including e.g. Google Apps, EVO, AG, Skype, Sakai and Plone and any other useful collaboration tools that emerge. International open source and commodity software and tools should be adopted wherever possible, where they meet the requisite quality

standards. It will also strongly encourage and advise researchers to adopt and participate in the development of international best practice in their discipline and research community whenever possible. A set of ARCS Internships, co-funded by ARCS and the host MARCS, will be established and offered toward the end of each Calendar year. They will be advertised widely and awarded to the most deserving through an open competition.

## **3.2. Services**

There is no charge for the use of generic ARCS Services by legitimate members of the Australian research community, but where special or dedicated services are required by a specific community or entity, then a mutually agreeable business arrangement between ARCS and the research community in question would be negotiated and entered into.

### **3.2.1. Collaboration Services**

The Collaboration Services group plans to have a full help desk operational by the beginning of 2008-09. This service will provide support, via telephone and email, to Australian researchers to help them use any of ARCS' eResearch services. The help desk will also operate as a single point of contact for new researchers to gain access to ARCS' services.

Further growth is envisioned in web based collaboration services operated by ARCS. It is expected tools, such as Sakai, Plone, and wikis, will provide a means for researchers to collaborate easily across institutional boundaries. ARCS will provide a managed service that is easy to set up and with low overheads, to facilitate this goal.

The Collaboration Services group will also be looking into providing support and guidance in a number of real-time collaboration tools, such as AccessGrid and EVO. An agreement is currently being negotiated with Caltech to provide EVO support, together with AARNet, to the Australian research and education sector. As part of this agreement ARCS will be operating the Tier 1 and 2 help desk support for EVO. Also planned is the furthering of the quality assurance of Australian AccessGrid nodes, to provide a better experience when using this technology.

### **3.2.2. Operational Services: Systems Services**

In the immediate future, the Systems Services Team expects to further 'harden' its existing production services and develop the other services towards production status. At the same time a close watch must be kept on developments elsewhere, particularly with respect to Authorisation and Authentication but also a large range of other technologies.

Much is expected of Grisu and its potential to allow a wide range of users with little tolerance for HPC jargon to use "The Grid" and several sites are eagerly awaiting the completion and deployment of the JMoIEditor, the Chemistry Portal project commenced during the APACGrid days and continuing as a Proto-NeAT project, and about to be adapted to suit a number of ARCS Member systems. Systems Support looks forward to its involvement in what is a growing area for ARCS, the Collaboration Services, where it has a role providing the underlying systems level support. Systems services will also be responsible for providing the underlying support for new services emerging from successful Proto-NeAT and NeAT Projects.

Longer term the ARCS Systems Services Team will be working closely with the emerging ARCS Authorisation Services Team and the AAF to migrate the existing infrastructure to the much more user-friendly approach we expect to see soon. Such migration will be well under way by the end of 2008 and by the middle of 2009 we expect to be totally dependant on this new approach.

There can be little doubt that the technology we currently use will continue to progress over the next year and ARCS will need to ensure its current high status internationally is maintained.

### **3.2.3. Operational Services: Data Services**

The data collaboration team will continue to develop the ARCS **Data Fabric** and to bring other planned services up to production level by packaging, deploying, testing and monitoring these services.

The ARCS Data Fabric is an easy to use way for researchers or groups of researchers (virtual organisations or VOs) to manage their data. End users will easily and transparently be able to obtain access to a minimal quantity of storage resources without needing to invest financially or technically. Supporting either the command line or GUI interfaces and providing a range of data movement and storage tools, the system will support Grid Security layers and Virtual Organisations. For more information see <http://www.arcs.org.au/datafabric>.

Work is progressing well for on the data fabric model. Early adopters should be involved in this activity by the middle of 2008 and by the first half of 2009 we would expect the service to be in routine use by researchers who consider it part of their normal workflow. The MARCS have committed through ARCS to providing mass storage to high-volume data users at cost for the disk and tape hardware. On top of this pervasive hardware data storage solution, ARCS will provide the overlying Data Fabric tools and services and will work with other data storage providers to bring their storage facilities into the fabric in a transparent way.

Based around physical resources contributed by ARCS Members and operated, as a service, by ARCS staff, the Data Fabric is expected to be produced at minimal cost and it will provide considerable benefit to end users. The intention is that some researchers will find such a service meets their immediate needs and others find it whets their appetite sufficiently for them to engage ARCS to develop further services for their specific use. From an efficiency point of view, ARCS would like end user groups to be able to use "out of the box" services but fully recognises that such an approach is unlikely to meet everyone's needs.

The provision of services will continue to key projects including the Australian Synchrotron, IMOS Remote Sensing and the Skymapper project. Experience from working with projects like these will continue to be applied to other projects as they arise.

Work will continue on a planned service to provide managed data transfers. This will be able to streamline data transfers by optimising the use of available file system and network resources. This file transfer service will be hooked into the proposed data fabric and other ARCS data services so that they also gain from its efficiencies.

Overall, the emphasis in the next year will be on developing and implementing the proposed data fabric. Once it goes into testing the data fabric will be modified as needed to suit the needs of the research community and will be flexible and accommodating with regard to implementing discipline and community specific tools and services.

### **3.2.4. Projects**

The primary purpose of the ARCS Projects activities is the delivery of the ARCS components of the NeAT Projects and to coordinate with ANDS and NeAT on this. This is the responsibility of the ARCS Projects Manager. From time to time ARCS may become involved in other non-NeAT projects and these projects would also be overseen by the ARCS Projects Manager. As each of ARCS and ANDS have received \$6 million to jointly fund NeAT projects over the NCRIS funding period and as it is expected that every Project

will have both ARCS and ANDS components, discussions have occurred and agreements have been made by ARCS and ANDS in order to facilitate these Projects.

NeAT is set up to develop integrative activities between PFC and the other NCRIS capabilities as well as the broader research community. Importantly, NeAT is the only source of such integrative funding. NeAT funding is budgeted on a continuous average expenditure of \$3million per annum, i.e., \$12 million in total over 4 years. NeAT has decided to allow for two rounds of decision-making: an initial set of projects to start in 2008-09 and a second set of projects to start in 2009-10. In addition, in the year 2007-08 extension activities from APAC were agreed in the interim and referred to as Proto-NeAT. The rationale given by NeAT for two rounds of major projects was:

- that not all decisions should be made at once, so that the experience in NeAT can be leveraged into new developments;
- that any significant development will require at least 2 years work because of the complexity of the technologies involved;
- that projects should be structured to complete within the funding period (complete by end of 2010-11).

**NeAT Project Governance:** ARCS and ANDS have discussed and jointly agreed on the management of NeAT Projects as follows:

- Each NeAT Project would have a NeAT Project Committee consisting of an ANDS representative (the Executive Director or delegate) and an ARCS representative (the Executive Director or delegate), representatives from any other institutions that would manage the enduring services provided by the NeAT Project, community nominated discipline representatives, a designated NeAT Project Manager (ex officio) and a prominent discipline leader as the NeAT Project Committee Chair. Where a suitable discipline Chair could not be found, the Chair would be either the ANDS or ARCS representative depending on whether the project was more ARCS or ANDS;
- Each NeAT Project would have a Project Manager who would be appointed by the relevant NeAT Project Committee;
- The Project Manager must be the person who manages the day to day work of the project and who will report directly to and be directed by the Project Committee;
- The governance structures of ANDS and ARCS would need to be satisfied with the Project Committee's management of the project in order to ensure the funds keep flowing, which provides the appropriate checks and balances and ensures accountability;
- If needed in the interim period pending ANDS formal establishment and only with the approval of the ANDS Project Management Committee (PMC), ARCS could be the vehicle for contractual arrangements with external entities should they be needed. Once ANDS is established, any contractual arrangements would typically include ANDS and ARCS as appropriate.
- Core responsibilities of each of the NeAT Project Committees include: overseeing and approving the design and implementation of an appropriate and relevant enduring service; and at the end of the Project identifying and establishing a Steering Committee of key stakeholders and service providers to manage this enduring service into the future and to take over from the NeAT Project Committee.
- ARCS and ANDS will jointly review each NeAT Project every three months using their standard processes and the NeAT Project Committees would review their project every six months with a written report from the NeAT Project Manager. NeAT would review all NeAT Projects annually in September, beginning 2009, as part of the established NeAT processes.
- ANDS and ARCS will pay their NeAT Project funds quarterly in arrears based on acceptable performance on a per EFT basis for each NeAT Project. The ANDS and ARCS quarterly reviews will be the trigger for either approving or withholding NeAT

funds for that quarter from a NeAT Project or a component of that project as appropriate.

**NeAT Project Cost Division:** ARCS and ANDS have discussed how the funding for NeAT projects should best be divided between ARCS and ANDS in the coming business year. The following have been agreed as principles to guide the funding allocation:

- All the proposed NeAT projects have both ANDS (data management) and ARCS (collaboration tools and services) related components; as such, both entities are interested in their success and care about their governance;
- All the proposed projects will have ANDS and ARCS representation on their steering committees;
- The precise nature of the projects (and the relative emphasis of ARCS or ANDS concerns within the projects) will not become clear until after the projects have commenced;
- The money in the NCRIS budgets for ANDS and ARCS should be seen as PfC money to be used for the greater good of NCRIS;
- ARCS and ANDS have developed a close and collaborative working relationship;
- ANDS and ARCS have agreed that the NeAT projects be funded in FY 2008/9 50% by ANDS and 50% by ARCS. This funding split will be reviewed for FY 2009/10 as part of the normal business planning cycles for both ARCS and ANDS.

Brief descriptions of the anticipated NeAT Projects to begin 1 July 2008 are included below. More detailed descriptions of these projects are provided as NeAT Project Business Plans in Appendix E. Each of these projects has been approved by NeAT, but will not proceed until a detailed Project Plan has been developed that is acceptable to the Project Committee and has been agreed by the ARCS Executive Committee and the ANDS Project Management Committee. The development of these plans is proceeding, with the aim of completing them before the end of May.

**NeAT Project - Marine and Climate Data Discovery and Access Project (MACDDAP):**

This project supports providers of marine and climate data sets by creating new and enhanced services, built on international standards and software, to more easily manage, translate and control these distributed digital repositories for the benefit of Australian researchers in a variety of disciplines. Data providers include NCRIS Integrated Marine Observing System (IMOS), CSIRO Marine Atmospheric Research, CSIRO Land and Water, the Bureau of Meteorology and the Australian Oceanographic Data Centre - Joint Facility. The main outcomes of this project will be a greater availability of marine and climate data in a wider range of standard protocols, including the Open Geospatial Consortium (OGC) standards for Web Map Services, Web Coverage Services and Web Feature Services, which will be integrated into OpeNDAP, the standard protocol used for accessing these data sets. Workflow tools will be created for processing existing data sets in order to generate standard metadata and enable the data to be accessed via the OGC services. New and enhanced services such as a Metadata Entry Search Tool, an OpeNDAP digital library metadata harvester, a Catalogue Exchange Service and an Aggregation Service will ensure that the distributed data sets will be discoverable, searchable, and increasingly conformable with standard vocabularies. Consequently, researchers will be able to collect and aggregate marine and climate data across disciplines for knowledge discovery.

**NeAT Project - Spatial Information Services Stack (SISS):** The SISS project will develop some of the component software components, services and functional capabilities needed to realise a spatial information data commons within Australia, conforming to Open Geospatial Consortium (OGC) and ISO standards. The spatial data infrastructures being developed by NCRIS capability areas such as AuScope and IMOS, as well as many other spatial data providers including the CSIRO Water Resources Observation Network, SEE Grid and the Western Australian Shared Land Information Platform (SLIP), will be able to interoperate,

using common software and services, and will establish a spatial information data commons that will be utilized by a variety of research communities. SISS will include complete applications or services comprising client interface (e.g. web portal or web service interface), middleware and data repository components. AuScope will use SISS to deploy: an OCG Catalog Service providing a web service interface to a registry with multiple registers to support both discovery of data and governance of standards associated with the spatial data commons (e.g. interoperability profiles, controlled vocabularies, data standards); a spatial information discovery portal; and SISS-based access to a number of spatial data repositories including GPS network station and observation logs, Seismic and Hyperspectral transects, Geological models, maps, and the Virtual drill core library.

**NeAT Project - Data Integration and Annotation Services in Biodiversity (DIAS-B):** This project consists of two related sub-projects that are of immediate benefit to NCRIS 5.2.3, the Atlas of Living Australia (ALA), and more widely to NCRIS 5.2 Integrated Biological Systems (IBS). However, the services developed for these sub-projects are expected to have much broader applicability, and may eventually be offered as general services by ANDS. This project will act as a test case for the application of these services to a particular discipline area. In order to provide integration across many and varied biodiversity data sets, the ALA needs to maintain a Metadata Repository to enable metadata registration and harvesting for all available digital resources of biodiversity information. The first sub-project relates to the development of such a repository and the processes and tools required for data providers and consumers to make use of the service. The second sub-project will develop authenticated data annotation services that will enable users or automated data analysis tools to provide feedback on data quality by annotating data records and resource metadata with comments on data quality and structured corrections. The original data providers will be able to connect to the service to retrieve and process annotations to their data, and provide their own comments.

**NeAT Project - Collaborative Integration and Annotation Services for Australian Literature Communities (AustLit):** The aim of this project is to provide services to address the needs of researchers involved in the study of Australian Literature. The main electronic resource for this community is AustLit, which carries comprehensive high quality metadata for book-length contributions to Australian literature, however access to full text and other data forms - images, short articles, etc. - is very limited. Currently large numbers of searches through many different interfaces across many data sets are needed. This project will develop data integration services and a web portal that federates search across several online databases and catalogues including AustLit, Scholarly Electronic Text and Image Service (SETIS), Australian National Bibliographic Database, Australian Digital Theses, Dictionary of Australian Artists Online, Australian Dictionary of Biography, National Library of Australia's PeopleAustralia and PictureAustralia and Project Gutenberg Australia. There is currently no electronic support for a core activity of literature researchers - to create commentaries and annotations, particularly collaboratively. The project will develop annotation services to enable communities of experts to collaboratively tag/annotate digital resources, including full text documents, with keywords, notes, comments, interpretations and links to related resources and to share these annotations with their community. Community members will also be able to create, publish, edit, search and retrieve OAI-ORE compliant Compound Objects that relate multiple resources from disparate databases through RDF named graphs.

**NeAT Project - ASSDA Services for e-Social Science (ASeSS):** The Australian Social Science Data Archive (ASSDA) provides online access to over 1000 social science data sets, including census data, surveys and opinion polls. Much of the high-value data that researchers want to access for secondary analysis has ethics and confidentiality requirements and these prevent the adoption of an open data commons approach. Thus a major challenge for ASSDA is the need to balance access and data analysis with assurance

of confidentiality, without requiring labour intensive processes. The project will develop web-based services for authenticated on-line data analysis, including cross-archive search, textual analysis, spatial data analysis, computational analysis, modelling and visualization. The data analysis tools will support the unlocking of otherwise inaccessible data sets of national significance, and enable groups of researchers to form ad-hoc groups to work together on a data set and problem, and to focus on analysis rather than development of analysis tools. The project will also develop data curation tools that will enable much lower cost acquisition of social science data.

**NeAT Project - A Data Fabric for Characterisation – Microscopy, Imaging, Neutron and X-ray Facilities (DataMINX):** This project will develop and deploy services that will enable researchers using NCRIS Characterization facilities, including the Australian Synchrotron, the OPAL neutron source at ANSTO, the distributed instruments of the Australian Microscopy and Microanalysis Facility (AMMRF), as well as institutional X-ray crystallography facilities, to make use of the ARCS data fabric, including distributed data storage facilities and data management and transfer services. The services will be based on the software and services that have been developed by the UK Science and Technology Facility Council (STFC) for its neutron and synchrotron facilities. Services to be implemented include metadata capture at source; generation of data in standard interchange formats such as imgCIF and NeXus/HDF; transfer of data from the instruments to managed data storage facilities hosted by the MARCS, VerSI or institutions; federated data management initially using Storage Resource Broker (SRB); a metadata catalog (ICAT) based on the STFC Scientific Metadata Model; and a web portal service, such as the STFC DataPortal, providing search and retrieval across facilities and platforms. Training material and documentation will be developed to help researchers make use of the services. The project will work to utilise the authentication and authorization services being developed by PfC, and investigate alternatives to SRB for future use.

**Other NeAT Projects:** Other potential NeAT projects are still under discussion by NeAT but have not yet been approved. These include but are not limited to:

- Bioinformatics;
- Virtual Observatory Web Services for Astronomy Data;
- Integration of humanities data sets using semantic web technologies.

Any new projects which go ahead in the Period will need to have a project proposal approved by NeAT, and a Business Plan and detailed Project Plan approved by ARCS and ANDS.

The astronomy project proposal is currently the most mature and a summary of this is given below.

**NeAT Project - Virtual Observatory Tools and Data Services for Astronomy (Aus-VO):** The astronomy community has developed international standards for metadata, data formats and web interfaces for searching and accessing astronomy data, through the International Virtual Observatory Association (IVOA), with Aus-VO as its Australian member. This project will enable Aus-VO, PfC and the MARCS to work together with the astronomy community and NCRIS 5.10 in order to develop best practices for the management, sharing and analysis of astronomical data, by implementing federated data repositories and IVOA-compliant data access services for several data sets of national and international significance that will be hosted by the MARCS. This includes data sets from existing optical and radio instruments (such as GAMA, WiggleZ, ATLAS, SUMSS), as well as new instruments funded by NCRIS 5.10 (SkyMapper and the SKA Pathfinder projects), each of which will produce hundreds of TBytes of data. The scientific return from implementing these services will be huge. The ease of use, and re-use, of archival data is a growing and

important requirement for modern astronomical research. A significant proportion of current astronomical research comes from re-analysis of existing data, the often-cited example being that of the Hubble Space Telescope, where around three times as many papers are written using archival data as from the original observations. The project will also have the benefit of significantly increasing eResearch practices and expertise within the astronomy community.

**Rate of NeAT Expenditure:** If we view NeAT funding as \$12 million of block funding across ARCS and ANDS, then:

- Proto-NeAT activities have expended approximately \$525k, which is approximately 4%;
- the 6 approved NeAT Projects in Appendix E expend approximately 50%;
- the remaining potential projects to be approved in the Period are estimated to expend approximately a further 15%;
- leaving approximately 31% of NeAT expenditure available for the second round which would become part of the 2009-10 Business Plans for ARCS and ANDS.

Any unspent funding, due to non-appointment or fractional delivery of effort delivery, can be returned to the second NeAT round. This expenditure profile will continue into the subsequent year and hence overspend the NeAT annual funding in that year, however NeAT Projects are being structured to complete within the three year funding period. Assuming similar funding of each significant NeAT Project in the second NeAT round, it follows that **only approximately five additional significant integrative projects can be implemented during the remainder of the NCRIS funding period.**

### 3.2.5. Authorisation Services

Subsequent to the email from Clare McLaughlin of DIISR on 6 Feb 2008 regarding the "ARCS Centre of Expertise in Authorisation", discussions have occurred between relevant parties and a governance structure and process have been agreed

There will be an Authorisation Funding Agreement that will be developed and agreed by DIISR and ARCS, which will then be signed by ARCS through its designated legal agent. This Authorisation Funding Agreement will then become an additional Schedule in the ARCS Collaboration Agreement that will determine how ARCS delivers Authorisation Services on behalf of Australia's research community. It will therefore form an additional and new service activity within ARCS and as such it will be subject to all of the usual ARCS governance structure and procedures.

We understand that INTERSECT will soon be formed in NSW and will sign on to the ARCS Collaboration Agreement as the NSW member replacing ac3. We also understand that Macquarie University will be a member of INTERSECT. Since there is significant expertise at Macquarie in the area of Authentication and Authorisation it is appropriate to initially concentrate a significant fraction of the Authorisation Services personnel there in the first year of operations. In the second and third years it is expected that some of the effort in this area will find itself moved to other places around the country. The primary purpose of the Authorisation Services effort is to spread expertise throughout ARCS, ANDS and other eResearch Service providers, to develop and implement important Authorisation exemplars using the Authentication services to be provided by AAF and to be a source of best practice and leading-edge knowledge and expertise.

In addition to the standard ARCS governance and management processes, a Steering Committee has been formed to provide advice to the ARCS Executive Committee and consists of Dr Rhys Francis (AeRIC Executive Director), Professor James Dalziel (Director of MELCOE) and the ARCS Executive Director. In addition, a larger Authorisation Working Group will develop the detailed technical plans for Authorisation Services and will include the

ARCS Executive Director and the ARCS Management Group, Professor James Dalziel, the CEO of INTERSECT and a number of ARCS representative(s) as determined by the ARCS Executive Committee.

### **3.3. Milestones**

It is important to establish and describe target milestones to indicate the expected dates of service availability, the enhancement of existing services and the deployment of new services over the Period:

#### **Before 31 July 2008:**

- ARCS announces that after an initial Establishment Period it is now formally 'Open for Business' and widely advertises and promotes this;
- Communication and Outreach Strategies finalised and initiated;
- ARCS Telephone Helpdesk fully deployed with the phone answered during all standard Australian business hours and with the email Helpdesk available for registering issues after hours;
- Establish an ARCS User Reference Group to provide advice and guidance on appropriate ARCS Services and new service directions;
- ARCS Data Fabric available to the whole research community with basic functionality and modest storage for free and for more significant storage needs at hardware cost;
- ARCS compute grid and GRISU will have their production services well-defined, in full production-mode operation and be well-documented on the ARCS web site with exemplars and use case studies;
- ARCS web site completed including clear and accessible explanations of the full suite of ARCS Production Services including exemplars and use cases;
- ARCS and EVOGH enter into a formal agreement for the licensing of EVO use in Australia, where legitimate use will encompass all activities and entities involved in furthering research and education in Australia;
- ARCS Authorisation Services Manager appointed and Authorisation Services activities underway;
- ARCS Collaboration Services Manager appointed and Collaboration Services undertakes a structured program of enthusiastic and proactive engagement with regional and discipline-based research communities and University IT Directors;
- ARCS engagement with New Zealand counterparts initiated and opportunities for cooperation explored;
- NeAT Project Committees established and NeAT Project Managers appointed and NeAT Projects underway;
- Each of ARCS Collaboration, Systems, Data and Authorisation Services Teams produce a written document describing their Quality Assurance and Change Management Procedures so that they can be placed on the ARCS web site and available for the research community and other Service Providers;
- Each NeAT Project Committee establishes its own Quality Assurance and Change Management Procedures and wherever possible these should be consistent with those of ANDS and ARCS;
- Explore with the NT Government and Charles Darwin University the possibility of establishing a MARCS in the Northern Territory to enable the delivery of ARCS Services and Capabilities into the region;
- Complete planning of the ARCS 1-day Workshop at the eResearch Australasia 2008 Conference in Melbourne.

#### **Before 1 October 2008:**

- Each ARCS Services Team establishes and targets a set of new user sub-disciplines and or communities and proactively establishes a cooperative relationship in order to explore and ultimately deliver a set of relevant services to that community;

- EVO@AU deployed and operational and Communities established;
- University roll-out of EVO@AU to Monash and interested first-round universities;
- Enhanced functionality of ARCS Data Fabric deployed;
- First ARCS Affiliate signs on;
- Formal engagement with CAUDIT established and first joint activities undertaken;
- Formal cooperative relationships with New Zealand counterparts established and activated.

**Before 1 January 2009:**

- Full functionality of ARCS Data Fabric deployed and available;
- Enhanced compute grid, GRISU and related services deployed, operating and documented;
- Possible formal establishment of a Northern Territory MARCS and signing on to the ARCS Collaboration Agreement;
- First ARCS Internships offered and accepted with internships underway;
- Second and third ARCS Affiliates sign on.

**Before 1 April 2009:**

- Review of performance of ARCS on a **service-by-service basis** using a suite of measures designed to establish: growth of user base by absolute number; distribution of user base by discipline/community; user satisfaction; user participation and engagement with ARCS.

### **3.4. Measures**

During the Period a number of steps will be taken to measure and improve the success and impact of the ARCS tools and services. These will include the monitoring of customer satisfaction as well as measuring the impact of services, the rate of new uptake of services and tools, the reach of the dissemination of knowledge and expertise into the research communities, the quality of services, the responsiveness of teams to new requirements and the implementation of change management practices to maintain production services at a professional level. Measures will include, but are not limited to:

- Quarterly analysis of service usage by the Services Teams with a focus on demonstrating impact into new research areas and communities;
- Annual user satisfaction surveys, carried out before April each year on a service by service basis;
- Written feedback from the ARCS User Reference Group;
- Invited submissions from a significant cross-section of the major research communities, including the NCRIS Capability Areas, as well as established discipline areas;
- Invited feedback from AeRIC, other national eResearch Service Providers, DIISR and other key stakeholder communities.

### **3.5. Communication and Outreach**

Communication and outreach strategies will be developed in detail before the end of July 2008, but will at least include:

- Professional design of a style and template for the ARCS web page;
- Uptake of key advertising strategies and opportunities;
- Part-time engagement of a professional publicist;
- Implementation and promotion of ARCS Internships;
- Wide promotion of the ARCS Helpdesk;
- Wide promotion of EVO and other readily appreciated ARCS Services;
- Engagement with cognate international bodies (e.g., Open Grid Forum, EGEE etc); development of ARCS exemplars and success stories; soliciting of testimonials re ARCS services; and

- Development of press releases and stories about ARCS, including those demonstrating the relevance of ARCS to the furthering of national priority goals and of leading to public good.

#### 4. Budget for the Period

Included below is a brief Budget Summary for the Period. The annual budget is presented in complete detail in Appendix A including names and locations of ARCS funded staff where known. It is largely self-explanatory and consists primarily of funds for ARCS personnel. The rates for EFTs for 2008-9 are \$100k pa for new and inexperienced recruits, \$120k pa for experienced personnel and \$170k pa for Managers. The full funds indicated for the EFTs will *only* be paid if the position was filled for the full year by an ARCS-approved person and in the manner outlined in the ARCS Position Description Template. For partial effort or partial occupancy of the position, the funding will be pro-rata.

<b>BUDGET SUMMARY</b>		
<b><u>Income:</u></b>	<b>DIISR funds (\$k)</b>	<b>In-kind (\$k)</b>
<b><i>Operations</i></b>		
Original ICI Operations Funds (\$14m over 4 yrs)	3500	
Collaboration Services Funds (\$0.5m over 3 yrs)	166.67	
Authorisation Services Funds (\$2m over 3 yrs)	666.67	
Interest (estimate)	240	
<b><i>Operations Income Sub-Total</i></b>	<b><i>4573.34</i></b>	
<b><i>NeAT</i></b>		
Original ICI NeAT Funds (\$6m over 4 yrs)	1500	
ICI NeAT Carry Forward from 2007-08 (estimate)	975	
<b><i>NeAT Income Sub-Total</i></b>	<b><i>2475</i></b>	
<b>Total known cash income</b>	<b>7048.34</b>	
<b><u>Expenditure:</u></b>		
<b><i>Operations</i></b>		
Executive, Administration & Lead Agent	388.5	
Non-personnel	220	
Operational Services (includes Systems services & Data Services)	2159	
Projects Manager (0.5)	85	
Collaboration Services	1070	
Authorisation Services	650	
<b><i>Operations Expenditure Sub-Total</i></b>	<b><i>4572.5</i></b>	<b><i>2800</i></b>
<b><i>NeAT</i></b>		
NeAT Projects (\$1,225k from 6 projects in Appendix E)	1225	2520
NeAT Projects (additional to be approved) – estimated cost	975	1950
<b><i>NeAT Expenditure Sub-Total</i></b>	<b><i>2200</i></b>	<b><i>4470</i></b>
<b>Total estimated expenditure</b>	<b>6,772.5</b>	<b>7,270</b>

In the non-NeAT area there was a slight underspend in 2007-8 because arrangements between CSIRO and ARCS regarding an acceptable position description were not finalized until Q2 2008. The funds unspent there will be used as carry forward into years 2009-10 and 2010-11 as a way of increasing the funding per EFT in operations if and when needed. All

carry-forward from ICI funding, for example, must and will be used to deliver ICI services during the NCRIS period.

Of the available \$1.5 million of NeAT funding in 2007-08, Proto-NeAT is expected to have expended approximately \$525k by the end of that first year. This leaves approximately \$975k of ARCS NeAT funding, which has been shown as carry-forward of NeAT income into the Period. As is evident from the Budget Summary Table, there is an anticipated carry forward of NeAT funding into 2009-10 of approximately \$2475k - \$2200k = \$275k. The Operations funds are essentially all exhausted including the estimated interest for the year, i.e., \$4573.34 - \$4572.5 = \$840.

**In-kind for Operations:** The Australian Research Council uses a standard in-kind multiplier of 1.25 based on salary plus on-costs in order to estimate university indirect in-kind contributions. The majority of ARCS staff are employed through universities. The typical salary plus on-costs for an ARCS staff member is approximately \$100k, (i.e., on costs are 29% at universities). Using the ARC formula this equates to a total average cost per EFT of \$100k + \$125k = \$225k of which \$120k is paid from ICI funds. This is equivalent to an in-kind contribution of order \$105k per EFT and since there are approximately 35 EFTs in Operations, this equates to an in-kind contribution of approximately \$3.5 million p.a. in addition to the ICI funding of \$4.57 million. To be conservative and because not all staff are housed at universities, it seems reasonable to discount the in-kind by of order 20% to give a conservative \$2.8 million. This comes, for example, in the form of office accommodation, infrastructure provided by the MARCS including computer room space, effort from non-ICI funded staff to assist ICI activities and projects and supervision of staff from non-ICI funded staff.

**In-kind for NeAT:** The 6 approved NeAT Projects in Appendix E total to an expenditure of \$2.45 million from NeAT funds and precisely half of this is to be sourced from ARCS NeAT funding, i.e., \$1.225 million. The NeAT funding is leveraging significant in-kind contributions from the MARCS and the discipline communities. These 6 projects leverage an in-kind contribution of more than 28 EFTs. Using the above university formula with a \$225k cost per EFT and then again being conservative and discounting by 20% gives an in-kind contribution of 5.04 million. Half of this is \$2.52 million and is leveraging the ARCS NeAT funds. **This corresponds to an in-kind to cash ratio of better than 2 to 1 for NeAT Projects.** It is very reassuring for NeAT to see this very significant level of co-investment. In the Budget Summary Table an in-kind ratio of 2 to 1 was assumed for the as yet undecided NeAT projects as well in order to make an in-kind estimate for all of ARCS NeAT funding in the Period.

## Appendix A: ARCS Budget for the Period 2008-9

### *Expected Income for the Period*

<u>Income Source</u>	<u>Income (\$k)</u>
<b>Platforms for Collaboration:</b>	
Original ICI Operations Funds (\$14m over 4 yrs)	3500
Original ICI NeAT Funds (\$6m over 4 yrs)	1500
ICI NeAT Carry Forward from 2007-08 (estimate)	975
Collaboration Services Funds (\$0.5m over 3 yrs)	166.67
Authorisation Services Funds (\$2m over 3 yrs)	666.67
ANDS Services Funds (not yet known)	0
Interest (estimate)	240
	<b>7048.34</b>
<b>Other NCRIS Capability Areas:</b>	
Australian Biosecurity Information Network (ABIN) (not yet known)	0
Other TBA (not yet known)	0
	<b>0</b>
<b>Non-NCRIS Income:</b>	
Other TBA (not yet known)	0
	<b>0</b>
<b>Total Income</b>	<b>7048.34</b>





## Appendix B: ARCS Position Description Template

Below is the current position description template used for ARCS positions. The highlighted sections are modified for each position to suit the needs of the position.

**Title:** Collaboration Services Team Member

**Name:** TBA

**Local Host:** TBA

1. **Position Summary:** The ARCS Collaboration Services Team is responsible for proactively providing friendly and helpful support for researchers who are using the ARCS Collaboration Service Tools. These services and tools include, but are not limited to, products such as Wiki, Drupal, Sakai, Plone, EVO, Access Grid and the like. The position requires a flexible individual with relevant technical expertise and experience, a willingness to work harmoniously in a team, and a commitment to develop and display a professional and helpful approach to clients in the research community at all times. The position may evolve over time and may be reassigned to one or more different ARCS teams in response to shifting demands on ARCS services.

### 2. REPORTING RELATIONSHIPS

This position is a full-time ARCS position and as for all ARCS activities the position will ultimately be responsible to the ARCS Executive Director. The ARCS Line Manager for this position will be the Collaboration Services Manager, who will determine the ARCS duties to be performed, manage the day-to-day activities and carry out appropriate performance reviews. While the initial primary role of the position is to be a part of the Collaboration Service Team under the direction of the Collaboration Service Manager, the Executive Director, or his/her delegate, reserves the right to assign additional duties and/or change the nature of the duties and the team membership over time to best meet the needs of ARCS. Should this occur a revised Position Description will be issued to reflect the revised duties and ARCS reporting lines.

Where it does not detract, interfere or conflict with ARCS assigned duties, the position may also deliver non-ARCS incidental services to the Local Host under the direction of the Local Host line manager. For incidental matters relating to the Local Host, such as security, time keeping, general employment conditions and leave this position reports directly to his/her supervisor at the Local Host. However, leave and working hours should be jointly agreed between the Local Host line manager and the relevant ARCS line Manager.

### 3. SPECIFIC ACCOUNTABILITIES

#### Area: ARCS Deliverables

- The ARCS mission, goals and deliverables are defined elsewhere and can be found on the ARCS web site [www.arcs.org.au](http://www.arcs.org.au);
- Work with NCRIS and other research groups to help define their Collaboration services, tools and application needs;
- As part of a team proactively support researchers in the uptake and use of these services and tools;
- As part of a team contribute to the development of these services and tools by identifying potential improvements and either developing those improvements or communicating the improvements to the appropriate developers (who may not be ARCS personnel);
- Investigate new technologies and services that may meet current or future Collaboration Services needs;
- Take a lead role in at least one of the identified collaboration services and/or tools;

- Contribute to the strategic direction setting and policy-making processes of ARCS by participating in appropriate ARCS working groups and forums.

**Area: Local Host accountabilities**

- Ensure that activities directed toward the Local Host are consistent with the Local Host rules and policies and that wherever possible are also consistent with the ARCS mission and deliverables;
- Ensure that Local Host Member's return on investment benefits of ARCS activities are appropriately reported and recorded;
- To contribute to the strategic direction setting and policy-making processes of the Local Host by participating in key working groups and forums.

**4. PERSONAL DEVELOPMENT**

To assume joint responsibility for career development by actively participating in various activities and producing a jointly agreed development plan identifying key areas which will assist in personal and professional development. These activities should be consistent with both the ARCS and Local Host Business Plans.

To maintain a high level of personal integrity when dealing with both organisations and people. To ensure the management, reporting staff and others are treated fairly and with respect.

**5. KEY SELECTION CRITERIA**

- Relevant tertiary qualification and/or substantial relevant experience;
- An ability to work closely with end users who may often lack particular technical skills and a strong commitment to proactively providing a friendly and high-quality service to all ARCS clients;
- Demonstrated technical ability or strong interest in one or more of the ARCS **Collaboration** services and/or tools;
- Excellent interpersonal skills necessary to work with team members both local and interstate as well as clients and other researchers in the broader Australian research community;
- Demonstrated ability to communicate clearly in both verbal and written form.

**6. OTHER RELEVANT INFORMATION**

The position may require work in other areas in the absence of colleagues, and/or outside normal business hours, as requested by appropriate Management.

It is agreed that the Local Host employs the person filling this role and

- Assigns the role to the line Manager of ARCS for ARCS activities;
- Will confer with ARCS if employment conditions change;
- Will confer with ARCS regarding any proposal to replace the incumbent;
- Will confer with ARCS if working hours becoming non-standard or are significantly changed;
- Will confer with ARCS before other significant non-ARCS duties are assigned by the Local Host;
- Appropriate ARCS line management will be notified if the incumbent is unavailable due to illness;
- Will consult with ARCS line management regarding holidays, training and to attend to urgent and time-consuming non-ARCS matters;
- Will inform and consult with ARCS if the incumbent is demonstratively unhappy in the role or appears to be unable to undertake the necessary tasks in an appropriate manner.

## Appendix C: ARCS Delegation of Powers

The ARCS Executive Committee has resolved that the following are the current delegation of powers:

Transaction	1 <sup>st</sup> Level	2 <sup>nd</sup> Level
Hire Permanent & Temporary Staff	ARCS Executive Director (with prior budget approval for permanent staff)	Respective Employer (Local Host)
Expenditure < \$50,000	ARCS Executive Director	—
Budgeted expenditure between \$50K and \$250K	ARCS Executive Director	—
Budgeted expenditure > \$250K	ARCS Executive Director	ARCS Chairman
Unbudgeted expenditure between \$50K and \$250K	ARCS Executive Director	ARCS Executive Committee
Invoices for payment, accompanied by an approved Purchase Order or other appropriate documentation	ARCS Executive Director	Lead Agent
Petty Cash Vouchers (Imprest \$500)	ARCS Coordinator	ARCS Executive Director
Courier Services	ARCS Coordinator	—
Entertainment Expenses up to the value of \$1K	ARCS Executive Director	—
Contracts for <b>ARCS</b> services up to the value of \$100K, using approved contract	ARCS Executive Director	—
Travel and Expenses — Domestic	ARCS Executive Director (with prior budget approval)	—
Travel and Expenses — International	ARCS Executive Director	ARCS Executive Committee

**Key:** 1<sup>st</sup> Level — party that initiates the transaction, and approves if no 2<sup>nd</sup> Level;  
2<sup>nd</sup> Level — party that approves the transaction

## Appendix D: Resolutions re Potential Conflicts of Interest

### Rationale and motivation:

The underlying premise is that the MARCS have joined together to form ARCS in order to cooperate on the delivery of *national eResearch support services* to the Australian research community in the most effective manner possible.

ARCS is about national teams delivering national eResearch support services to the Australian research community. ARCS is a subset of the sum of activities of the MARCS, just as ICI and Authorisation are two subsets of the activities of ARCS. It is therefore inevitable that at times the aspirations of one or more of the MARCS for new business opportunities may lead to conflict of interest with ARCS or one or more of the other MARCS.

Where core ARCS-related services are to be delivered to a national eResearch support effort, such as one of the NCRIS capability areas, the most natural mechanism of delivery is ARCS itself rather than one of more of the MARCS. An example is the delivery of services to ANDS by ARCS. Clearly, this is core ARCS business and indeed ARCS is explicitly mentioned in the ANDS documentation as the appropriate means of delivery of some key ANDS activities. Similarly, since ARCS is being funded to deliver Authorisation it would not be appropriate for one of more of the MARCS to compete in that area.

### Recommendations ratified by the ARCS Executive Committee:

1. *Definition of 'ARCS Activities'*: By definition ARCS activities are those subject to the governance procedures of ARCS and as such will involve the Executive Director and Executive Committee in both their development and their delivery. Any activities not matching this definition are not ARCS Activities and should not be represented as such. It is anticipated that most ARCS activities will be national eResearch support services.
2. *Definition of 'ARCS Agreements'*: An agreement is only an 'ARCS Agreement' when ARCS itself is a signatory on the agreement.
3. *Definition of 'ARCS Business'*: A business arrangement is only an 'ARCS Business' arrangement, where the flow of monies is through the ARCS accounts and hence subject to the usual ARCS governance arrangements. Typically ARCS Business would have an associated ARCS Agreement.
4. *An 'ARCS Register of Interests' is to be maintained by ARCS*: Where a conflict of interest has a possibility of occurring between the current or future activities of ARCS and the planned or ongoing activities of one or more of the MARCS, then the relevant MARCS will inform the Executive Director and the Executive Committee so that an appropriate entry can be made in the ARCS Register of Interests.
5. The MARCS agree that the ARCS is the preferred vehicle for the delivery of national e-Research support services and to work together to pursue opportunities for sustainability and growth of the ARCS.

It has further been agreed that communications from ARCS and representations by ARCS can only come either through the Chair of the Executive Committee or through the Executive Director.

## Appendix E: NeAT Project Plan Summaries

### ARCS-ANDS Agreed NeAT Governance Arrangements – (duplicated from the ARCS Business Plan for convenience):

**NeAT Project Governance:** ARCS and ANDS have discussed and jointly agreed on the management of NeAT Projects as follows:

- Each NeAT Project would have a NeAT Project Committee consisting of an ANDS representative (the Executive Director or delegate) and an ARCS representative (the Executive Director or delegate), representatives from any other institutions that would manage the enduring services provided by the NeAT Project, community nominated discipline representatives, a designated NeAT Project Manager (ex officio) and a prominent discipline leader as the NeAT Project Committee Chair. Where a suitable discipline Chair could not be found, the Chair would be either the ANDS or ARCS representative depending on whether the project was more ARCS or ANDS;
- Each NeAT Project would have a Project Manager who would be appointed by the relevant NeAT Project Committee;
- The Project Manager must be the person who manages the day to day work of the project and who will report directly to and be directed by the Project Committee;
- The governance structures of ANDS and ARCS would need to be satisfied with the Project Committee's management of the project in order to ensure the funds keep flowing, which provides the appropriate checks and balances and ensures accountability;
- If needed in the interim period pending ANDS formal establishment and only with the approval of the ANDS Project Management Committee (PMC), ARCS could be the vehicle for contractual arrangements with external entities should they be needed. Once ANDS is established, any contractual arrangements would typically include ANDS and ARCS as appropriate.
- Core responsibilities of each of the NeAT Project Committees include: overseeing and approving the design and implementation of an appropriate and relevant enduring service; and at the end of the Project identifying and establishing a Steering Committee of key stakeholders and service providers to manage this enduring service into the future and to take over from the NeAT Project Committee.
- ARCS and ANDS will jointly review each NeAT Project every three months using their standard processes and the NeAT Project Committees would review their project every six months with a written report from the NeAT Project Manager. NeAT would review all NeAT Projects annually in September, beginning 2009, as part of the established NeAT processes.
- ANDS and ARCS will pay their NeAT Project funds quarterly in arrears based on acceptable performance on a per EFT basis for each NeAT Project. The ANDS and ARCS quarterly reviews will be the trigger for either approving or withholding NeAT funds for that quarter from a NeAT Project or a component of that project as appropriate.

**NeAT Project Cost Division:** ARCS and ANDS have discussed how the funding for NeAT projects should best be divided between ARCS and ANDS in the coming business year. The following have been agreed as principles to guide the funding allocation:

- All the proposed NeAT projects have both ANDS (data management) and ARCS (collaboration tools and services) related components; as such, both entities are interested in their success and care about their governance;
- All the proposed projects will have ANDS and ARCS representation on their steering committees;

- The precise nature of the projects (and the relative emphasis of ARCS or ANDS concerns within the projects) will not become clear until after the projects have commenced;
- The money in the NCRIS budgets for ANDS and ARCS should be seen as Pfc money to be used for the greater good of NCRIS;
- ARCS and ANDS have developed a close and collaborative working relationship;
- ANDS and ARCS have agreed that the NeAT projects be funded in FY 2008/9 50% by ANDS and 50% by ARCS. This funding split will be reviewed for FY 2009/10 as part of the normal business planning cycles for both ARCS and ANDS.

## NeAT Business Plan Component

### Spatial Information Services Stack (SISS)

#### 1. Service Description

##### ***1.1. Description of a research community and the eResearch service need***

The complexity of data integration is rapidly increasing (more data sources and more combinations of interest) and traditional data integration methods have become untenable as the difficulty exceeds the available human resources. There is increasing need for researchers to access data and services based on open standards for interoperability.

Spatial data infrastructure is being developed by NCRIS capability areas such as AuScope, IMOS and Atlas of Living Australia, as well as many other spatial data providers including the CSIRO Water Resources Observation Network, SEE Grid and the Western Australian Shared Land Information Platform (SLIP). This infrastructure should be interoperable, with common tools and services conforming to Open Geospatial Consortium (OGC) and ISO standards, in order to establish a spatial information data commons that will be utilized by a variety of research communities.

##### ***1.2. Description of the proposed service solution and how it meets that need***

The SISS project will develop some of the software components, services and functional capabilities needed to realise a spatial information data commons within Australia, conforming to Open Geospatial Consortium (OGC) and ISO standards. SISS will include complete applications or services comprising client interface (e.g. web portal or web service interface), middleware and data repository components. While the software and services that are developed will be generic, the project will focus primarily on addressing the needs of the geosciences community, particularly AuScope.

AuScope will use SISS to deploy an OGC Catalog Service providing a web service interface to a registry with multiple registers to support both discovery of data and governance of standards associated with the spatial data commons (e.g. interoperability profiles, controlled vocabularies, data standards), and an associated spatial information discovery portal. These will be hosted by ARCS.

AuScope will provide SISS-based access to a number of spatial data repositories including GPS network station and observation logs, Seismic and Hyperspectral transects, Geological models, maps, and the Virtual drill core library. The members of AuScope holding the relevant data will sustain these services into the future.

The SISS project will build a skill base able to work with holders of spatial data to deploy and operate relevant data servers and OGC compliant services and a consequent increase in the number and variety of spatial data sets made available through common access mechanisms.

#### 2. Benefits and proposed measures

### **2.1. Benefits to the user community and associated measures**

The SISS should reduce the per unit cost of data publishing and access, particularly for research requiring data from a diversity of domains. Many research activities require access to data held by government agencies and this stack will facilitate improved access to spatial data of interest. Research addressing larger problems (systems, cross-disciplinary research) with more meaningful questions and answers should be possible.

Monitoring of the use of SISS-enabled portals and data access services will measure the volume and variety of data that is accessible through standardised mechanisms, and increases in data use, as well as access patterns.

### **2.2. Benefits to ANDS or ARCS (or other provider) and associated measures**

ANDS will improve its understanding of methods for accessing loosely coupled spatial data. There will be an increased use of spatial data hosted by government agencies (e.g. geological surveys), the MARCS (e.g. CSIRO), and other spatial data providers.

### **2.3. Expected flow-on benefits to others**

Use of OGC standards and standard software components for serving spatial data will progress the creation of a national spatial data commons. Improved data search and discovery services, and access to data using widely-used OGC standards, will enable the use of data by a range of user communities.

Communities outside AuScope and geosciences that are expected to benefit from the outcomes of the project are very broad, but include IMOS, Atlas of Living Australia, Terrestrial Ecosystem Research Network, Biosecurity, water resources (CSIRO WRON, AWRIS, e-Water CRC), Australian Bureau of Statistics, social sciences.

## **3. Resources and commitments**

### **3.1. Resources provided by the user community**

AuScope will provide 5 EFT and CSIRO e-SIM will provide 1 EFT of effort working on SISS or SISS-related development effort. An estimated 3 EFT of this effort would be working directly towards the goals of this project.

### **3.2. Resources provided by ANDS or ARCS**

NeAT funding of \$400K p.a. for 2 years and nominally \$200K for a third year. The third year amount will depend on the outcome of project reviews and available NeAT funds.

It is anticipated that the NeAT funding for this project would be used primarily to hire software developers through IVEC. However this will be decided by the Project Committee and specified in the Project Plan.

Some ARCS Operations effort will be utilized to deploy SISS at MARCS hosting spatial data sets. A rough estimate of this effort is a total of 0.5 EFT p.a. across all the MARCS.

### **3.3. Resources provided by others**

Some effort in deploying SISS by a variety of organizations hosting spatial data, including government agencies such as state Geological Surveys, and MARCS such as CSIRO.

**Total project resources and commitments are summarized in the following table.**

	Cash Y1	EFT Y1	Cash Y2	EFT Y2	Cash Y3	EFT Y3	Cash Total	EFT Total
User community (AusScope, CSIRO)		3		3		3		9
ANDS	200K		200K		100K		1.0M	1.5
ARCS	200K	0.5	200K	0.5	100K	0.5		

## 4. Governance

### 4.1. Governance processes to be applied to the project

- The ARCS/ANDS agreed governance mechanism for NeAT projects, defined in the ARCS and ANDS Business Plans.
- The Project Committee will meet quarterly via phone and/or agreed electronic medium.

### 4.2. Quality assurance processes to be used by or applied to the project

The ARCS/ANDS agreed arrangements will apply.

### 4.3. List of names against key governance and project management roles

Project Committee:

- ARCS Executive Director, Professor Anthony Williams, or nominee
- ANDS Executive Director, or nominee
- CEO Auscope, Scott McTaggart (Chair)
- General Manager, Office of Spatial Data Management, Ben Searle
- Geological Survey of Victoria, Alan Willocks
- Geoscience Australia representative (to be announced)
- CSIRO Land and Water Chief, Neil McKenzie, or nominee

The Project Manager will be appointed by the Project Committee and specified in the Project Plan.

## 5. Project Summary

### 5.1. Deliverables / Milestones

#### July 2008 - Dec 2008

- ARCS Hosting OGC Catalog Services and Discovery Portal
- Auscope hosting OGC Catalog Service and Discovery Portal
- Auscope GPS data WFS service deployed
- Geological Survey of Victoria GeoSciML testbed collaboration deployed

#### Jan 2009 - Dec 2009

- ARCS SISS support service fully established
- CSIRO Minerals Down Under Flagship deploys laterite geochemical data service (Western Yilgarn)

- Auscope NVCL WFS service deployed
- CSIRO Minerals Down Under Flagship deploys airborne hyperspectral data service
- CSIRO Minerals Down Under Flagship deploys thermodynamic data service
- CSIRO Minerals Down Under Flagship deploys Northern Yilgarn hydro-geochemistry

**Jan 2010 - Dec 2010**

- Auscope Virtual Rock Laboratory and Tsunami workflows utilise registry and information service infrastructure for service discovery and data management support
- Auscope deploys WCS for geophysics imagery with large data set support

**Jan 2011 - June 2011**

- Auscope Earth Model and portal service infrastructure fully established and using the service stack
- Broader adoption well underway

## **5.2. Overall risk assessment**

The major risks and their mitigation are:

- Project complexity. Project characteristics are similar to a previous successful project. The proposed project leader has a track record of successfully managing projects of this size.
- Stakeholders cannot agree on a common set of requirements. Communications involve stakeholders in a group situation (the Project Committee and Reference Group) as well as one-on-one. All stakeholders are demonstrating high levels of commitment to the project.
- Unable to secure appropriate and skilled staff for the required work. Can use secondments from partner organisations, and contractors to augment the development of the services.
- Technology development is highly challenging. There is a parallel approach in the project plan to mitigate greater than normal development risk. There is a contingency plan to recover significant value should the project not fully succeed. All stakeholders are fully aware of and accept the chances for project success.

## **5.3. Review points**

Quarterly reviews by ANDS and ARCS, six monthly written reports from the Project Manager to the Project Committee, and a yearly review each September (starting 2009) by NeAT.

## NeAT Business Plan Component

### Marine and Climate Data Discovery and Access Project (MACDDAP)

#### 1. Service Description

##### ***1.1. Description of a research community and the eResearch service need***

Researchers across marine and climate communities need to better use their combined data resources, as well as instruments to observe physical and biological properties around Australia. Basic services are being established through IMOS. This project supports providers of marine and climate data sets, by creating efficient services built on international standards and software, to more easily manage, translate, and control these distributed digital repositories for the benefit of Australian researchers.

##### ***1.2. Description of the proposed service solution and how it meets that need***

The underlying services that are to be provided by this project are :

Access Services: to enhance availability of marine and climate data in a wider range of standard protocols, including Open Geospatial Consortium (OGC) standards, integrated into OpenDAP, the standard protocol used for accessing many of these data sets.

Discovery Services: a Metadata Entry Search Tool, an OpenDAP digital library metadata harvester, a Catalogue Exchange Service and an Aggregation Service will ensure that the distributed data sets are increasingly discoverable and conformable with standard vocabularies.

Support Services: Workflow tools that process existing data sets in order to generate standard metadata and enable the data to be accessed via the OGC services.

The services will be delivered through the MARCS, particularly TPAC and CSIRO, but also the Bureau of Meteorology and others.

#### 2. Benefits and proposed measures

##### ***2.1. Benefits to the user community and associated measures***

Users will be able to more easily discover marine and climate data, and to access it using OGC standard web service interfaces. Development of automated workflows will reduce the cost of providing access to data and metadata using standard formats and interfaces.

Monitoring of portal use will measure increases in data use, and increases in the use of OGC standard interfaces, as well as access patterns. The impact of the services will be formally measured before and after automated services are created.

##### ***2.2. Benefits to ANDS or ARCS (or other provider) and associated measures***

ANDS will be able to understand and measure the extent to which marine and climate researchers access IMOS. In particular, much of this data is geo-referenced, ANDS will improve its understanding of methods for accessing loosely coupled spatial data.

There will be an increased use of data hosted by the MARCS and accessed through marine and climate data portals hosted and maintained by the MARCS.

### **2.3. Expected flow-on benefits to others**

Use of OGC standards for serving spatial data will progress the creation of a national spatial data commons. Improved data search and discovery services, and increased access to data using widely-used OGC standards and OpeNDAP protocol, will enable the use of data by a wider range of user communities.

Many important international data providers utilize the OPeNDAP and GeoNetwork software that will be enhanced by the work on this project, and hence will also benefit from the outcomes of this project.

## **3. Resources and commitments**

### **3.1. Resources provided by the user community**

IMOS will provide \$10K per annum cash and 3.2 EFT (2.5 from eMii, 0.7 from remote sensing stream).

### **3.2. Resources provided by ANDS or ARCS**

NeAT funding of \$400K p.a. for 2 years and nominally \$200K for a third year. The third year amount will depend on the outcome of project reviews and available NeAT funds.

Approx 1.5 EFT of ARCS Operations effort will assist AAF support, monitoring support for developed services, and deployment and hardening of services at the MARCS.

### **3.3. Resources provided by others**

1.25 EFT from Bureau of Meteorology and CSIRO, 1.2 EFT from TPAC.

**Total project resources and commitments are summarized in the following table.**

	Cash Y1	EFT Y1	Cash Y2	EFT Y2	Cash Y3	EFT Y3	Cash Total	EFT Total
User community (IMOS)	10K	3.2	10K	3.2	10K	3.2	30K	16.95
Other (TPAC, BoM, CSIRO)		2.45		2.45		2.45		
ANDS	200K		200K		100K		1M	4.5
ARCS	200K	1.5	200K	1.5	100K	1.5		

## **4. Governance**

### **4.1. Governance processes to be applied to the project**

- The ARCS/ANDS agreed governance mechanism for NeAT projects, defined in the ARCS and ANDS Business Plans.
- The Project Committee will meet quarterly via phone and/or agreed electronic medium.

### **4.2. Quality assurance processes to be used by or applied to the project**

The ARCS/ANDS agreed arrangements will apply.

### **4.3. List of names against key governance and project management roles**

Project Committee:

- ARCS Executive Director, Professor Anthony Williams, or nominee
- ANDS Executive Director, or nominee
- IMOS Executive Director, Professor Gary Meyers (Chair)
- eMii Director, Professor Roger Proctor (or Ms Kate Roberts)
- ARC NESS Convenor (Prof. Andy Pitman)
- Prof. Nathan Bindoff

The Project Manager will be appointed by the Project Committee and specified in the Project Plan.

## **5. Project Summary**

### **5.1. Deliverables / Milestones**

Major deliverables for the project are:

- Discovery Services
  - MEST enhancements (GEONETWORK, OPeNDAP)
  - TPAC Digital Library Portal enhancements (data registration, geospatial coordinate aware, gridftp, WCS and WMS service aware, simple visualisation)
  - Enhanced OPeNDAP harvester (geospatially aware, geospatial coordinate aware, gridftp, WCS and WMS service aware, simple visualisation)
  - Aggregation Service for remote sensing
- Access Services
  - OPeNDAP data delivered as WMS, WCS and trialled with AAF
  - OPeNDAP enhancements (authentication, administration, remote management, and data handler enhancements, server-side functions)
- Support Services
  - Translation Service (web based workflow for OPeNDAP to ISO19115 standards)
- Deployment of the above services at the providers nominated in 1.2.
- The enhancements of the software, community profiles, translation tools and related workflows in the proposed services will be delivered back to the international community, through their respective mechanisms or provided on Sourceforge where appropriate

Milestones in the first year of the project are:

- Development of OGC WMS services for OPeNDAP THREDDS server.
- Deployment of updated OPeNDAP servers at sites in Australia
- Prototype Aggregation Service
- Data conversion tools and workflows
- Improvements to TPAC Data Portal
  - Allow searches on spatial information
  - Improve search performance
  - Provide data visualization in portal
  - Services for registering data sets and associated metadata with the portal

Further details of deliverables and milestones are described in the MACDDAP Project Plan.

### **5.2. Overall risk assessment**

The major risks and their mitigation are:

- Unable to secure appropriate and skilled staff for the required work. Mitigation is the use of secondments from partner organisations, and contractors to augment the development of the services.
- Projects that depend on each other and cause time delays. This has been mitigated by minimising co-dependencies between projects.
- Uptake of the new services (e.g. translation service for pre-existing data sets so that they can be advertised in the eMii marine catalogue.). The mitigation strategy is built in the plan by allowing funds for the active interaction with staff that hold data and to develop specialised plug-ins for the data providers' particular environment.

### **5.3. Review points**

Quarterly reviews by ANDS and ARCS, six monthly written reports from the Project Manager to the Project Committee, and a yearly review each September (starting 2009) by NeAT.

# NeAT Business Plan Component

## Data Integration and Annotation Services in Biodiversity (DIAS-B)

### 1. Service Description

#### ***1.1. Description of a research community and the eResearch service need***

The Atlas of Living Australia (ALA) needs to support integration of a wide range of different types of biodiversity data – taxonomic data (e.g. taxon names and synonyms), specimen and observation data, species descriptions and associated images, diagnostic keys, genomic data, etc – from many different data providers. The user community for the ALA is very broad, encompassing taxonomists, botanists, zoologists, environmental scientists, land-use and conservation planners, and biosecurity officers. The Australian Centre for Plant Functional Genomics will be a specific user of plant phenomic data mediated through the project.

In order to provide discovery and interoperability across many and varied biodiversity data sets, the ALA requires needs best practices for metadata management, including adoption of relevant vocabularies and ontologies, and the ability to map between different metadata models. A Metadata Repository is required to enable metadata registration and harvesting for all available digital resources of biodiversity information.

The quality and consistency of the ALA data is crucial for its use. There is a need for an authenticated annotation service that will allow users or automated data analysis tools to provide information to users and feedback to data providers by annotating data records and resource metadata with comments on data quality and suggested corrections.

#### ***1.2. Description of the proposed service solution and how it meets that need***

Data quality services:

- Annotation service allowing human and machine users to store and retrieve annotations relating to any data record within the ALA to record possible errors.
- Reporting service that alerts data provider/owners of possible quality issue.

Data integration services:

- Catalogue of mandated and supported data standards, vocabularies, ontologies for use within the ALA.
- Metadata repository and metadata registration software for registration of all Australian biological data resources and for relating data sets to supported vocabularies and ontologies.
- Search interfaces (including web service interfaces) to search the metadata repository using terms from the supported ontologies.

Together, these services will enable data providers and researchers to actively participate in the creation and use of the Atlas of Living Australia.

These services will initially be hosted by ALA, however if these services can be made more generic they may be hosted by ARCS on behalf of ANDS.

### 2. Benefits and proposed measures

#### ***2.1. Benefits to the user community and associated measures***

Easier and faster integration of new data sources into ALA, measured by:

- Level of provision of metadata to metadata repository

- Number of data sets accessible via ALA
- Number of records accessible through ALA.
- Usage of ALA

Improved data discovery and federated search, measured by:

- Direct use of metadata services from the repository by other networks and repositories
- Percentage of data resources with metadata entries including references to ontology terms
- Extent of discovery through ontologies

Improvement of data quality via use of annotation services, measured by:

- Number and range of annotations in annotation database
- Number of responses from data providers
- Direct use of services (UI and web services) for providing annotations (other than through the ALA portal UI and ALA data validation tools services)
- Direct use of services (UI and web services) for accessing annotations
- Number of data records for which annotations have led to corrections in source data

Improved level of user experience, measured by:

- Independent reviews to be contracted at the end of 2008-2009 and at the end of 2010-2011 to document the experience of key target user groups and to compare the state of ALA infrastructure with other national biodiversity information platforms.
- An online survey tool to allow users to document their experience in using the ALA infrastructure. This survey tool will be continuously available as a data capture method. This survey will explicitly determine success in using data quality annotation.
- Analysis of web logs to determine whether users are guided to relevant information.

The Project Plan and the ALA Business Plan will specify some quantitative goals for these metrics for each year of the project.

## ***2.2. Benefits to ANDS or ARCS (or other provider) and associated measures***

There are two key benefits being sought. Firstly, an ability to integrate several different data sets with different schemata or ontologies, enabling researchers to find things despite having different knowledge lenses – this will be measured using the surveys described above. Secondly, we wish to understand how well an annotation service supports improvement of data quality – it will be measured using the survey described above.

## ***2.3. Expected flow-on benefits to others***

As articulated above, this affects all disciplines where commentary on other work is important. More specifically, the services developed within this project would also be of benefit to:

- NCRIS 5.12 Marine Sciences and Climate
- Social Sciences, ASSDA, AustLit
- NCRIS 5.3 Microscopy and Microanalysis
- NCRIS 5.8 Bio-security
- NCRIS 5.11 Terrestrial Ecosystem Research Network

# **3. Resources and commitments**

## ***3.1. Resources provided by the user community***

The Atlas of Living Australia will have approximately 5 EFTs working on software development related to this project. The Australian Biological Resource Survey (ABRS) and the Australian Museum also have existing developers whose products will be contributing directly to the development of the Atlas.

### **3.2. Resources provided by ANDS or ARCS**

NeAT funding of \$400K p.a. for 2 years and nominally \$200K for a third year. The actual amount for the third year will be dependent on the outcome of project reviews and available NeAT funds.

It is anticipated that the NeAT funding for this project would be used primarily to hire software developers at sites with relevant expertise, possibly including CSIRO, SAPAC, ANU and UQ. This will be decided by the Project Committee and specified in the Project Plan.

### **3.3. Resources provided by others**

There will be significant related international effort in standards development by the Taxonomic Data Working Group (TDWG) and open source software development effort from members of TDWG, GBIF and EoL (the Encyclopedia of Life). It is expected that some of these standards and software will be utilized in the NeAT project. This effort is difficult to quantify and not included here.

Total project resources and commitments are summarized in the following table.

	Y1		Y2		Y3		Total	
	Cash	EFT	Cash	EFT	Cash	EFT	Cash	EFT
ALA		5		5		5		15
ANDS	200K		200K		100K	1	1.0M	1
ARCS	200K		200K		100K			

## **4. Governance**

### **4.1. Governance processes to be applied to the project**

- The ARCS/ANDS agreed governance mechanism for NeAT projects, defined in the ARCS and ANDS Business Plans.
- The Project Committee will meet quarterly via phone and/or agreed electronic medium.

### **4.2. Quality assurance processes to be used by or applied to the project**

The ARCS/ANDS agreed arrangements will apply.

### **4.3. List of names against key governance and project management roles**

Project Committee:

- ARCS Executive Director, Professor Anthony Williams, or nominee
- ANDS Executive Director, or nominee
- Director of the Atlas of Living Australia, Donald Hobern
- Chair to be appointed.

The Project Manager will be appointed by the Project Committee and specified in the Project Plan.

## **5. Project Summary**

### **5.1. Deliverables / Milestones**

#### Metadata Repository Activities - Year 1:

- Review of metadata management and requirements in related international biodiversity informatics projects (particularly GBIF and EOL)
- Review metadata standards and ontologies in use within relevant Australian and international projects, and mappings between them.
- Review available software options for a metadata repository.

#### Metadata Repository Activities - Years 2-3:

- Contribute to the development of the TDWG core ontology.
- Establish standards for the use of unique identifiers for data resources and data items.
- Develop user interfaces and web services for primary registration of data resources and for configuration of OAI-PMH harvesting.
- Develop user interfaces and web services for search and selection of data resources via ontology terms as well as free-text search.
- Develop alternative output metadata formats (based on review of metadata standards above).
- Investigate how to integrate outputs from the Annotation Service into metadata management.

#### Annotation Service Activities - Year 1:

- Investigate requirements for annotation services in other NCRIS capabilities and in ANDS.
- Investigate existing collaborative annotation systems and select the most appropriate solution.
- Investigate how to integrate it with the Metadata Repository and other components in the ALA system.

#### Annotation Service Activities - Years 2-3:

- Develop an appropriate user interface that may need to be customised for structured annotation of different types of data.
- Test automated annotation of records by error-checking tools.
- Develop interfaces for management of obsolete annotations (e.g. after data record has been corrected for errors) and for threaded annotations (e.g. data provider responses to user comments)
- Test and refine the interface with a variety of users.
- Provide support for AAF authentication.

## 5.2. Overall risk assessment

Risk	Mitigations
Take up is slow	<ul style="list-style-type: none"> <li>• Measure take up</li> <li>• Project leader to be responsible for identifying projects and collaborations to provide and consume metadata and annotations</li> </ul>
Poor software	<ul style="list-style-type: none"> <li>• Good people/place – specifically oversight of experts from UQ, SAPAC and ANU</li> </ul>
Other approaches are more attractive	<ul style="list-style-type: none"> <li>• Keep watching and be adaptive in project development to make best use of developing standards and practices</li> </ul>

**5.3. *Review points***

Quarterly reviews by ANDS and ARCS, six monthly written reports from the Project Manager to the Project Committee, and a yearly review each September (starting 2009) by NeAT.

# NeAT Business Plan Component

## Collaborative Integration and Annotation Services for Australian Literature Communities (AustLit) Project

### 1. Service Description

#### ***1.1. Description of a research community and the eResearch service need***

AustLit currently provides a Web portal to arguably the most important collection of information for scholarly research into Australian literature and print culture. The Association for the Study of Australian Literature (ASAL) is the peak body for this research community. The current collection primarily supports bibliographic records with some full-text and limited image, audio or video content. The aim is to expand the AustLit collection to include more comprehensive access to full-text works as well as related content (images, recordings, reviews, critiques etc). In parallel, the search interface will also need to be enhanced to enable support for metadata, full-text, empirical and multimedia searching of external sources via a single web portal. Currently a large number of searches through many different interfaces across many data sets is required. As well, there is no electronic support for a core activity of literature researchers – to create and share collaborative commentaries via document markup or annotation tools. Finally, researchers would like to be able to encapsulate related artefacts into e-learning objects that they can share and exchange with other researchers and educators.

#### ***1.2. Description of the proposed service solution and how it meets that need***

The above needs will be met by the creation of several new services:

1. Data integration, search and reporting services
  - OAI-PMH metadata harvesting services that periodically harvest metadata from the specified external repositories and aggregate the metadata with the AustLit metadata repository
  - Federated Search Portal supporting metadata and full-text search across multiple data sets, and search, retrieval and presentation of records, text and images
2. Collaborative annotation services
  - Secure annotation creation, editing and attachment services
  - Annotation presentation, browse and search services
  - Annotation storage and servers (with authenticated access control)
  - Annotation harvesting and aggregation services
3. OAI-ORE compliant compound object authoring, editing and publishing services.

AustLit-specific services will be hosted and operated by the UQ Library, which currently hosts the AustLit server. The more generic services that may be deployed in disciplines other than AustLit will be operated initially by UQ's eResearch group and in the long term by UQ's IT Services and/or ARCS.

### 2. Benefits and proposed measures

#### ***2.1. Benefits to the user community and associated measures***

This project will widen the scope of AustLit, providing more comprehensive access to Australian literature resources. The project will also promote and expedite the adoption of eResearch practices by humanities researchers in Australia. Success will be measured by monitoring the number of site visits, searches and references to AustLit, both by

researchers, but also by web-links such as from Wikipedia, and scholarly works. The project will also enable groups of researchers to discover all material relevant to say “the visit of D. H. Lawrence to Australia”, or “the battle of Lawson and Paterson”, to create a collective annotated bibliography that can be explored and added to in a way that the current best practice – a book – cannot. A core research method of most of the humanities, law and the social sciences will be “e-enabled”. Results will be measured by monitoring usage of the services, the creation of new resources, and interviews with researchers on tool usefulness. The AustLit advisory board is derived from active researchers and teachers in the field and they will be involved in the testing and feedback cycle as the project develops.

## **2.2. Benefits to ANDS or ARCS (or other provider) and associated measures**

The ability to search seamlessly across both the metadata and the underlying data, particularly text, is an important one for many research areas. Much of the data access is envisaged to be by metadata, yet this is not the case for Google, say – this work will demonstrate how we bridge this boundary. The project will also provide an exemplar of data integration, with a federated search across multiple distributed databases. The only relevant measure here is that researchers use this new search facility and keep doing so. The ability to annotate data and other objects, and groups of objects, is important to many fields, but especially to the humanities. We need to demonstrate that this can be e-enabled – at the moment search is electronic, but not this aspect of the “work” of the humanities researcher. We will measure this principally by use and output. The ability to easily author compound objects in a standardized format that can be shared and re-used will greatly facilitate the use of AustLit and other cross-disciplinary content within eLearning resources.

## **2.3. Expected flow-on benefits to others**

As articulated above, this affects all disciplines where commentary on other work is important. More specifically, the services developed within this project would also be of benefit to:

- NCRIS 5.12 Marine Sciences and Climate
- Social Sciences, ASSDA
- Atlas of Living Australia, Terrestrial Ecosystem Research Network (TERN),
- NCRIS 5.3 Microscopy and Microanalysis and NCRIS 5.8 Bio-security

## **3. Resources and commitments**

### **3.1. Resources provided by the user community**

In-kind from AustLit, including Kerry Kilner, and other data providers.

### **3.2. Resources provided by ANDS or ARCS**

NeAT funding of \$250K p.a. for 2 years and nominally \$250K for a third year. The actual amount for the 3rd year will be dependent on the outcome of project reviews and available NeAT funds.

It is anticipated that the NeAT funding for this project would be used primarily to hire software developers at UQ. This will be decided by the Project Committee and specified in the Project Plan.

### **3.3. Resources provided by others**

In-kind from UQ Library for hosting the servers and services, and UQ eResearch group, including Jane Hunter. \$35K p.a. for 3 years from University of Queensland. \$35K p.a. in-kind for 3 years from QCIF.

Total project resources and commitments are summarized in the following table.

	Y1		Y2		Y3		Total	
	Cash	EFT	Cash	EFT	Cash	EFT	Cash	EFT
User community resources		0.5		0.5		0.5	210K	3.0
Others (UQ, QCIF)	70K	0.5	70K	0.5	70K	0.5		
ANDS	125K		125K		125K		750k	0
ARCS	125K		125K		125K			

## 4. Governance

### 4.1. Governance processes to be applied to the project

- The ARCS/ANDS agreed governance mechanism for NeAT projects, defined in the ARCS and ANDS Business Plans.
- The Project Committee will meet quarterly via phone and/or agreed electronic medium.
- A Reference Group will also be established for consultation.

### 4.2. Quality assurance processes to be used by or applied to the project

The ARCS/ANDS agreed arrangements will apply.

### 4.3. List of names against key governance and project management roles

Project Committee:

- ARCS Executive Director, Professor Anthony Williams, or nominee
- ANDS Executive Director, or nominee
- ASAL President, Elizabeth McMahon, or nominee (Chair)
- AustLit Executive Manager, Kerry Kilner
- UQ eResearch, Professor Jane Hunter

The Project Manager will be appointed by the Project Committee and specified in the Project Plan.

## 5. Project Summary

### 5.1. Deliverables / Milestones

Year 1: Full text search over both metadata and content with a variety of presentation formats.

Integrate 5 key databases, build and support federated search interface: AustLit, SETIS, Australian Dictionary of Biography Online, Dictionary of Australian Artists Online, and National Library services - National Bibliographic Database (NBD), People Australia, Picture Australia, PANDORA, and Australian Digital Theses DB.

Year 2: Collaborative Annotation creation, editing and publishing services

Year 3: Compound Object authoring and publishing services (encapsulate objects related by theme)

More detailed deliverables and milestones will be specified in the Project Plan.

### **5.2. Overall risk assessment**

The key risk of the project is that the tools are not used by Australian literature researchers - this is mitigated by the close engagement with the community and rapid delivery of beta versions for user acceptance.

### **5.3. Review points**

Quarterly reviews by ANDS and ARCS, six monthly written reports from the Project Manager to the Project Committee, and a yearly review each September (starting 2009) by NeAT.

## NeAT Business Plan Component

### ASSDA Services for e-Social-Science (ASeSS)

#### 1. Service Description

##### **1.1. Description of a research community and the eResearch service need**

The social science and humanities communities have need of a national e-Research infrastructure that enables researchers to explore their research data using a data commons approach that preserves confidentiality. This project involves setting up the first cohesive Australian e-Research service for the social sciences. The project aims to build a national model that incorporates closely related research domains that forms the foundation for an e-Social Science Virtual Organisation (SSVO).

The Australian Social Science Data Archive (ASSDA) collects and preserves computer-readable data relating to social, political and economic affairs and makes the data available for further analysis. The primary *user* community of ASSDA is Australian academic researchers, doctoral candidates, postgraduates and undergraduates in the social sciences and humanities. However, the services offered by ASSDA are structured in a fashion that also allows researchers and staff from Australian public sector agencies, non-government organisations, the media and, occasionally, the private sector to access 'public use' data sets. Much of the high-value data that researchers would like to access for analysis has ethics and confidentiality requirements. Thus a major challenge presented to ASSDA is the need to balance access with assurance of confidentiality, without requiring labour intensive processes.

##### **1.2. Description of the proposed service solution and how it meets that need**

The ASeSS project will develop a suite of services to support the SSVO. These will include:

- Data curation software that supports reliable data ingestion, and sets up the appropriate access controls.
- Search tools that support authenticated discovery across archives
- Integration of a suite of analytic tools that enable easy use by social scientists (the tools themselves will be developed outside this project)
- Integration of a suite of visualisation tools, particularly spatially oriented and temporally oriented

The SSVO will need to integrate a wide variety of services, and will need strong authentication and authorization before access and use (in most cases). Providing online access to statistical data analysis tools provides additional value to the data archives and also supports the 'unlocking' of data sets of national significance that cannot otherwise be made accessible due to privacy constraints.

#### 2. Benefits and proposed measures

##### **2.1. Benefits to the user community and associated measures**

ASeSS should reduce the cost of data publishing, access and analysis. Social scientists should have access to a suite of tools that are easy to use as standard desktop publishing

tools. The key measure will be the extent to which social scientists use this toolset, and the data that the toolset supports. The cost of data acquisition is currently high, so it will be important to ensure that efficient data ingestion and curation occurs.

Monitoring of the use of ASeSS will measure the volume and variety of data that is accessible using web monitoring tools and user surveys.

## **2.2. Benefits to ANDS or ARCS (or other provider) and associated measures**

ANDS will improve its understanding of methods for data acquisition and access over a wider variety of data types than is typically held in numerically oriented databases. As well, tools that support the exploration of a combination of textually and numerically oriented data will be developed. Social science data is an important element of the data commons so it is important that it is well represented in the data commons. ARCS is charged with supporting Virtual Organisations, and ASeSS will enable better understanding of VO's with strong authentication and access control. It should be a good testbed for AAF.

## **2.3. Expected flow-on benefits to others**

This tool set should be generalisable and thus be of benefit to many user communities that need analytic and visualisation tools for a wide variety of data types. The VO infrastructure will be of value in any research community where there is a need to collaborate within a restricted environment.

# **3. Resources and commitments**

## **3.1. Resources provided by the user community**

ASSDA will provide 6 EFT.

## **3.2. Resources provided by ANDS or ARCS**

NeAT funding of \$400K p.a. for 2 years and nominally \$200K for a third year. The third year amount will depend on the outcome of project reviews and available NeAT funds.

It is anticipated that the NeAT funding for this project would be used primarily to hire software

developers through ASSDA. However this will be decided by the Project Committee and specified in the Project Plan.

Some ARCS Operations effort may be utilized to deploy ASeSS at MARCS and other institutions hosting social science data sets.

## **3.3. Resources provided by others**

This has not yet been established.

**Total project resources and commitments are summarized in the following table.**

	Cash Y1	EFT Y1	Cash Y2	EFT Y2	Cash Y3	EFT Y3	Cash Total	EFT Total
User community (ASSDA)		6		6		6		18
ANDS	200K		200K		100K		1.0M	
ARCS	200K		200K		100K			

## 4. Governance

### 4.1. Governance processes to be applied to the project

- The ARCS/ANDS agreed governance mechanism for NeAT projects, defined in the ARCS and ANDS Business Plans.
- The Project Committee will meet quarterly via phone and/or agreed electronic medium.

### 4.2. Quality assurance processes to be used by or applied to the project

The ARCS/ANDS agreed arrangements will apply.

### 4.3. List of names against key governance and project management roles

Project Committee:

- ARCS Executive Director, Professor Anthony Williams, or nominee
- ANDS Executive Director, or nominee
- ASSDA Director, Dr. Deborah Mitchell
- Research community representatives, to be determined

The Project Manager will be appointed by the Project Committee and specified in the Project Plan.

## 5. Project Summary

### 5.1. Deliverables / Milestones

Period	Major milestones
Dec 2008	Demonstration version of GIS data viz web tool Demonstration version of longitudinal data analysis web tool
June 2009	First release of new ASSDA VO web site Demonstration of first text-based VO service deployment Demonstration of cross-archive data search between two major archives Historical Census and Colonial Data Archive (HCCDA) availability
Dec 2009	Time Series & Panel Archive and Qualitative Data Archive available Text based analysis tool demonstration for Qualitative Data Archive Search available on UR data archive
June 2010	Search available for Qualitative Data and HCCDA on VO web site Longitudinal data analysis web service available on VO web site Data exchange service for Time Series & Panel Data and Qualitative Data Q/A component of curation software for Qualitative data Production version of web-based UR archive curation service on VO web site
Dec 2010	Time Series& Panel Archive services available on VO web site

	Q/A component of Indigenous data archive curation service Search available on Electoral Database on VO web site
June 2011	Release production version of web-based Qualitative data archive curation service on VO web site Search available over Indigenous archive Release production version of Indigenous data archive curation service on VO web site Generalised version of GIS service on UR data available on VO web site

## **5.2. Overall risk assessment**

- The major risks and their mitigation have not yet been completed.

## **5.3. Review points**

Quarterly reviews by ANDS and ARCS, six monthly written reports by the Project Manager to the Project Committee, and a yearly review each September (starting 2009) by NeAT.

## **NeAT Business Plan Component**

### **A Data Fabric for Characterisation – Microscopy, Imaging, Neutron and X-ray Facilities (DataMINX)**

#### **1. Service Description**

##### ***1.1. Description of a research community and the eResearch service need***

Thousands of researchers use molecular and material characterisation facilities, such as those funded by NCRIS 5.3, including the neutron facilities at ANSTO, the Australian Synchrotron, institutional X-ray diffraction facilities, and the Australian Microscopy and Microanalysis Facility (AMMRF), which is a network of major institutional facilities and linked labs of national significance. These facilities currently do not provide the capability for researchers to easily and reliably transfer experimental data from the facility to remote data storage or a data repository, with automated capture and storage of associated metadata, and the ability to easily provide authenticated sharing of the data to colleagues or to publish the data in a way that it is easily discoverable and accessible by any researcher. Ideally these services should be uniform across the different facilities and leverage the national data fabric provided by ARCS and the data storage and data repositories provided by the MARCS and other institutions. A significant number of researchers who use Australian characterisation facilities also use international facilities, hence interoperability and linkage with international facilities is highly desirable.

##### ***1.2. Description of the proposed service solution and how it meets that need***

The project will develop a “data fabric” of data services and infrastructure to support the above user and facility requirements. The underlying services that will be developed by this project will include: automated capture of data and metadata from instruments and transfer to a data repository; conversion of data and metadata to standard formats; a federated data repository; a searchable metadata catalog; fast data transfer and download; and a web portal to discover and access data, with authorized data access.

The services will build on the ARCS data fabric delivered by the ARCS Data Services team, utilizing data storage provided by the MARCS and other organizations (e.g. VerSI).

It is expected that the services will be based as much as possible on existing software and services, such as those developed by the UK Science and Technology Facilities Council (STFC) for use in the ISIS neutron facility and Diamond synchrotron in the UK, and by the DART and ARCHER projects. These services are based on Storage Resource Broker (SRB) and the ICAT metadata catalog, which implements the STFC Scientific Metadata Model.

The system will initially be targeted at AMMRF, Neutron, Synchrotron and X-ray facilities, but the aim is to extend it to support the Imaging component of NCRIS 5.3 once a working system is deployed and in use.

The detailed specification of the project is still under development and will be specified in the Project Plan.

## **2. Benefits and proposed measures**

### **2.1. Benefits to the user community and associated measures**

An important aim of the project is to move users away from personal, ad-hoc data management and data sharing practices, to the use of use of automated data management services, secure data access and sharing, and institutionally supported long-term archiving of data and associated metadata. The provision of these services will mark a significant change in current community practice, protecting and preserving research outputs, and making it easier for research collaborations to share data and to utilize results from multiple distributed characterisation facilities. This will improve the level of uptake of sound eResearch practice as individual researchers and research groups utilise both the collaborative aspects of the data grid as well as its archival and data re-use/publication aspects.

In addition, characterisation data sources will no longer be viewed in isolation. Instruments will become linked through data grid technologies and services, and this will make it easier for researchers to undertake studies that rely on more than one type of instrument.

The project will monitor the number of experiments and users utilizing the new services, and the amount of data being archived and shared. User community views will be measured by an annual survey.

### **2.2. Benefits to ANDS or ARCS (or other provider) and associated measures**

The proposed project provides some large user communities and different use cases for the ARCS data fabric and associated authentication services.

ANDS will be able to understand and measure the extent to which a large distributed research community can make use of services for capturing data and metadata from various scientific instruments and moving it to federated data archives, and the use of authenticated data sharing. The proposed solution can also be used as a solution for institutions that need an integrated data management solution.

Through this project, ANDS will understand how well metadata can be captured with data directly from an instrument that supports both the immediate use, and subsequent uses of characterisation data. As well, ANDS will be informed about metadata transformation for use in an authenticated federated data archives.

Measures will be as described in section 2.1.

### **2.3. Expected flow-on benefits to others**

This project will develop a data management system that will support data and metadata capture from instruments, data transfer to federated data repositories, authenticated data access, with an associated metadata catalog that supports a general and extensible scientific metadata standard. It therefore has the potential to be much more broadly applicable.

## **3. Resources and commitments**

### **3.1. Resources provided by the user community**

ANSTO will provide approx 1 EFT of in-kind effort, particularly input into the design and testing of the system, as well as funding for storage and hosting of their data at ac3.

Australian Synchrotron (AS) will provide some in-kind for IT support.

MMSN and Uni of Sydney will provide some in-kind programming effort and input into the design and testing of the system. Expected to be able to provide approximately 1 EFT in the first year, subject to review after that.

AMMRF has funding for 2 new EFTs to support this project, and in-kind contributions from existing staff at the AMMRF facilities for service development and deployment and assistance with managing the AMMRF-focussed effort will be at least another 1.5 EFT.

### **3.2. Resources provided by ANDS or ARCS**

NeAT funding of \$600K p.a. for 2 years and nominally \$300K for a third year. The third year amount will depend on the outcome of project reviews and available NeAT funds.

It is expected that the NeAT funds would be used to hire software engineers to assist in customizing existing software and services for particular facilities and user requirements, developing new software and services where required, and working with staff at the experimental facilities to deploy, test and document the services and to assist users in making use of the services. At least one NeAT-funded person would need to be closely associated with each facility – the Australian Synchrotron in VIC, ANSTO and X-ray facilities in NSW, and the AMMRF nodes, which are located in QLD, NSW, ANU, WA and SA.

Significant effort from ARCS Operations will also be provided, through the ARCS Data Services team. This is estimated to be at least 3 EFT. Some effort from the ARCS authorization services project is expected to focus on the requirements of this project.

### **3.3. Resources provided by others**

The MARCS have agreed to provide storage capacity at hardware cost.

VeRSI is expected to provide at least 1 EFT.

INTERSECT is expected to be able to provide approximately 1 FTE for the first year, with follow-on to be reviewed in the final quarter of the first year.

The ARCHER project is undertaking work that is closely related to this project. Some ARCHER effort in 2008H2 could be aligned with this project and software developed by ARCHER could be used. An assessment of ARCHER developments will be made as part of the project.

The project will leverage software developed by the UK Science and Technology Facilities Council (STFC) that is used for the ISIS neutron facility and Diamond synchrotron in the UK. The STFC, ISIS and Diamond e-Science groups are willing to collaborate on the project.

**Total project resources and commitments are summarized in the following table.**

	Cash Y1	EFT Y1	Cash Y2	EFT Y2	Cash Y3	EFT Y3	Cash Total	EFT Total
User community (ANSTO, AS, AMMRF, MMSN)		5.5		5.5		5.5		22.5
Other (VeRSI, INTERSECT)		2		2		2		
ANDS	300K		300K		150K		1.5M	9
ARCS	300K	3	300K	3	150K	3		

## **4. Governance**

### **4.1. Governance processes to be applied to the project**

- The ARCS/ANDS agreed governance mechanism for NeAT projects, defined in the ARCS and ANDS Business Plans.
- The Project Committee will meet quarterly via phone and/or agreed electronic medium.

### **4.2. Quality assurance processes to be used by or applied to the project**

The ARCS/ANDS agreed arrangements will apply.

### **4.3. List of names against key governance and project management roles**

The Project Committee is still being decided. Proposed members include:

- ARCS Executive Director, Professor Anthony Williams, or nominee
- ANDS Executive Director, or nominee
- Dr Allan Jones, Chair of AMMRF eResearch Committee
- Nick Hauser, ANSTO
- Richard Farnsworth, Australian Synchrotron
- Dr Peter Turner, Manager of the ARC Molecular and Materials Structure Network (MMSN)
- Director of VeRSI, or nominee
- Prof Jane Hunter, UQ

The Project Manager will be appointed by the Project Committee and specified in the Project Plan.

## **5. Project Summary**

### **5.1. Deliverables / Milestones**

The deliverables for the project are still under discussion. Some proposed deliverables are presented below but these have not yet been agreed. Details of deliverables and milestones will be specified in the Project Plan.

Year 1

- Investigation of possible solutions for the federated data repositories, metadata catalog and associated web portal, including work done by STFC, ARCHER, VeRSI and GRANI project
- Investigate how the proposed system would interface to the ARCS data fabric and data storage and repositories at the MARCS and other organizations such as VeRSI and universities
- Develop a design for the system building on existing software and the ARCS data fabric
- Design and prototyping of software and services for data and metadata capture from the various instruments and transfer to data repositories
- Development of software for converting data and metadata to standard formats
- Investigation of data sharing, and authorization and authentication mechanisms and development of prototype AAA services based on AAF

- Investigate approaches for automating workflows for data and metadata capture, conversion to standard formats and schemas, other data processing, and ingestion into data repositories.
- Deployment of services for data transfer from ANSTO and some Australian Synchrotron beamlines, AMMRF facilities and X-ray labs to MARCS and other organizations and institutions which have appropriate storage support.
- Functional data repositories for ANSTO and some AMMRF facilities.

#### Year 2

- Authentication and authorization mechanisms integrated with AAF and ARCS services.
- Development of documentation and researcher training material, some researcher training programs run
- Some data processing workflow services implemented
- A federated data repository and prototype collaborative work environment for AMMRF facilities
- Widening the deployment of the system to additional appropriate facilities, instruments, beamlines and data types not addressed in Year 1
- Extension of services to interface other repository services such as provided by SPECTRA, e-Crystals and TARDIS.
- Ongoing investigation of interoperability issues and technology developments (e.g. iRODS)

### **5.2. Overall risk assessment**

To be completed in the Project Plan.

### **5.3. Review points**

Quarterly reviews by ANDS and ARCS, six monthly written reports from the Project Manager to the Project Committee, and a yearly review each September (starting 2009) by NeAT.