Australian National Data Service (ANDS)

National Collaborative Research Infrastructure Strategy 2013 (NCRIS 2013)

Progress Report 2

For the period 1 January 2014 to 30 June 2014

31st July 2014

Contents

1 Introduction

This progress report outlines the current progress against the intended actions as per the submitted NCRIS 2013 Project Plan.

The project plan described the intended actions to be carried out in the period September 2013 until June 2015 using NCRIS 2013 funding. The project plan is derived from an overall two year plan that describes the plan for ANDS activity over that whole time using NCRIS, EIF, DWF, CRIS and NCRIS 2013 funds. The programs, or components, are the same as have been agreed for the CRIS funding.

The programs, or components, are:

- **National Collections**, ensuring that the formation and curation of research data collections, together with associated services, is maintained;
- **National Services**, ensuring that the research data services, both technical and advisory, that have been established under the ANDS programs to date are maintained;
- **Institutional Engagement**, ensuring that institutional research data infrastructure is maintained, and continues to develop, using institutional funds, in a cost effective and nationally coherent manner; and
- **International Collaboration**, ensuring that the research data collaboration internationally is maintained, principally through the established international partnerships, most notably the Research Data Alliance.

These four programs, taken together, will enable ANDS to continue its data partnerships with:

- Our research institutional partners – focused on the Universities and the PFROs.
- Our research data providing partners – this includes NCRIS capabilities, national data generating facilities, public and private sector data providers, international data providers, and the public.
- Our research data infrastructure partners – this includes RDSI, NeCTAR, NCI and Pawsey, regional data service providers, and institutional eResearch partners.
- Our international partners, notably through RESEARCH DATA ALLIANCE, but including the US, the EC, New Zealand, U.K, the Dutch, amongst others.

The particular focus of this phase of ANDS activity is research data content rather than research data services and infrastructure. We will ensure that the infrastructure that has been set up through the Australian Research Data Commons is maintained and operated, however there will be increased emphasis on making data visible, available and reusable. This will be carried out using the internal resources of ANDS as well as partnerships funded by three engagements:

- *Major Open Data Collections engagements* – partnering with research institutions to promote collections that are of enduring significance to the institutions and that support their research strategy and research partnerships strategy. The result of these engagements would be that at least 26 internationally significant open data collections are supported by the institutions and exploit Australia's research data infrastructure.
- *The Access to Data for Research Infrastructure Capability Areas engagement* will enable ANDS to partner and fund the eight capability areas that are data intensive, and already have a strong ANDS partnership, to provide continued support for data collection publication, to enable services that are being generated through the NeCTAR Virtual Laboratories to be connected, provide long term access to data, reliable access to collections via rich descriptions, and connections to international collections through Research Data Australia.

- **The Maintaining Connections engagements** will ensure that connections between institutional metadata stores and RDSI node data registries enable institutions to have a comprehensive view of their research data assets, as well as ensuring connections are made between the collections held on RDSI nodes, their descriptions in Research Data Australia, and the developing data services, such as ones being developed through the NeCTAR Virtual Laboratories.

2 National Collections Program

The scope of this program includes collections of government, institutional, discipline and national facility data.

Two ANDS programs, Public Sector Data and National Collections, were combined to form the National Collections program. To date Public Sector Data has engaged with over nineteen agencies to raise the profile of data management and expose their collections. It has resulted in over thirty-four thousand collections and datasets being published through Research Data Australia. National Collections has promulgated awareness of the value of a collections-focussed approach to data and gained some traction in acceptance of the value of distributed national collections. A proof of concept for these distributed collections has been developed and demonstrated with positive results. The merging of these programs is indicative of the importance public data collections play as an input to research. There is increased value in a collections approach when combined with the delivery of storage solutions and services in partnership with other organisations.

The National Collections program will be responsible for the Access to Data for Research Infrastructure Capability Areas engagement, which will enable ANDS to partner and fund the eight capability area that are data intensive, and already have a strong ANDS partnership, to provide continued support for data collection publication, to enable services that are being generated through the NeCTAR Virtual Laboratories to be connected, provide long term access to data, reliable access to collections via rich descriptions, and connections to international collections through Research Data Australia.

The Access to Data for Research Infrastructure Capability Areas engagements will provide up to $200,000 in funds and strong ANDS engagement with each of the following areas that have been identified in the *2011 Strategic Roadmap for Australian Research Infrastructure*:
- Marine Environment
- Terrestrial Systems
- Solid Earth
- Urban Settlements
- Astronomy
- Population Health Research Platforms
- Integrated Biological Discovery
- Climate Change

This will enable ANDS to continue its engagement that has informed the thinking of the Research Data Infrastructure Committee, and enable further engagement with IMOS, TERN, ALA, AuScope, AURIN, AA, PHRN, BPA, APPN, NIF, and AMMRF.

2.1 Highlights, Difficulties and Breakthroughs

Nil to report.

2.2 Activities Undertaken

Although no Access to Data for Research Infrastructure Capability Areas activities were undertaken during the reporting period using NCRIS 2013 funds, a number of preparatory and planning activities were undertaken using a combination of NCRIS and EIF funding, and are thus reported in the NCRIS and EIF final reports.

2.3 Performance against agreed Key Performance Indicators

Nil to report.

3 National Services Program

The scope of this program includes collection, registration, publication, discovery, and advisory services.

In its first four years of operation ANDS has established a national registry of research data, a national discovery portal, and a data publication identifier service. These are nationally focused, value-adding services which are more efficient to run as a single instance for the sector. For example ANDS has joined the global DataCite infrastructure consortium and now provides data Digital Object Identifiers to the whole Australian research sector. Through the EIF and NCRIS infrastructure, many research organisations have established data archives and metadata provider infrastructure to feed the ANDS national registry, which aims to provide a national cross-discipline view of the data outputs from (and inputs to) Australian research.

The NCRIS project has enabled ANDS to establish a capacity to support, advise and exchange information about data infrastructure establishment at Australian research institutions. As well it provides targeted assistance for those using ANDS national services. ANDS has also established an expertise and advice service around questions of national policy promoting reuse of research data.

The National Services program will be responsible for the Maintaining Connections engagements. It will ensure that connections between institutional metadata stores and RDSI node data registries enable institutions to have a comprehensive view of their research data assets. In addition it will ensure connections between the collections held on RDSI nodes, their descriptions, in Research Data Australia and the developing data services, such as ones being developed through the NeCTAR Virtual Laboratories are connected. These engagements will enable ANDS, RDSI, and NeCTAR to fund Nodes to ensure that there is an easy way to use NeCTAR funded services over data described in Research Data Australia and held on an RDSI node. There will be connectivity between the data and the services, and the data will be maintained and visible. $1,000,000 is allocated to this engagement with the specific approach to coordinating and optimising ANDS, RDSI, and NeCTAR investments to be determined jointly between the projects, the service providers, and the service users.

3.1 Highlights, Difficulties and Breakthroughs

Nil to report.

3.2 Activities Undertaken

Although no Maintaining Connections activities were undertaken during the reporting period using NCRIS 2013 funds, a number of preparatory and planning activities were undertaken using a combination of NCRIS and EIF funding, and are thus reported in the NCRIS and EIF final reports.

3.3 Performance against agreed Key Performance Indicators

Nil to report.

4 Institutional Engagement Program

The program will ensure that institutional research data infrastructure is maintained and operated at all of Australia's major research organisations to ensure an continuing coherent Australian Research Data Commons, and that the research data assets of the institutions are made as valuable as possible.

This program will maintain activity funded in previous years under the Seeding the Commons, Data Capture, Metadata Stores and Applications programs. These programs built a fabric for data management with a view to increasing the scope of the data commons; improved data capture and management across the research sector; funded the establishment of infrastructure in institutions for depositing data and metadata into well-managed stores; and funded the creation of a number of demonstrations of the value of bringing data together to answer new questions.

The final year of NCRIS and EIF funding has enabled new data focused partnerships that will assist the organisations involved to achieve their research data ambitions through the continuation of a series of targeted engagements. These engagements are supported by a service offering that includes advice, tools and training and can be offered within a set of categories such as management, publication citation, re-use.

Major Open Data Collections

Under NCRIS 2013, the Institutional Engagement Program will be responsible for the Major Open Data Collections (MODC) engagement. The last report indicated that ANDS would provide $200,000 to 26 institutions that had installed data capture and/or institutional metadata stores to establish long-term open access to one or more collections of strategic significance to the institution. This number has now been revised and funding will be offered to the following 24 institutions;

- Monash University
- La Trobe University
- University of Melbourne
- Deakin University
- RMIT University
- Swinburne University
- Flinders University
- University of Adelaide
- University of South Australia
- University of NSW
- University of Sydney
- University of Technology
- University of Wollongong
- University of Western Sydney
- Macquarie University
- Australian National University
- CSIRO
- University of Tasmania
- James Cook University
- Queensland University of Technology
- Griffith University
- University of Queensland
- University of Western Australia
- Curtin University

The number of institutions has been reduced due to a new approach taken by ANDS regarding how it would engage with the Regional University Network (see section 4.1). The offer of funding has been declined by the University of Newcastle. This institution was in support of the invitation to take part in an MODC engagement, but was not in a position to take up the offer due to internal restructuring and resourcing.

The Major Open Data collections are expected to have the following characteristics:
- Substantial collections of significant data already exist
- There is a demonstrated research need for the data beyond the scope of the institution
- The data should be well described and well linked having richly connected data collections and sub-collections
- Open access to the data will be provided, although allowing for the possible requirement for registration
- The data collection must be discoverable through appropriate means including Research Data
- Australia as well as institutional, international and discipline specific portals
- The institution can demonstrate its enduring commitment to this data collection

ANDS has engaged with senior representatives of each institution to establish their desire to establish a Major Open Data Collection. Some institutions choose to establish more than one such collection. And although not a requirement of the funding, institutions were encouraged to invest in-kind support on top of the ANDS financial support (which is limited to $200,000), as it is envisaged that these collections will a require long-term commitment by the institution. Provision of ANDS expertise into the project was also part of the offer, which was formalised via a project charter which outlined agreed division of responsibilities and activities between the institution and ANDS, and which formed part of the contract with the institution.

This $5.2 million dollar engagement strategy is intended to ensure that Australian research institutional data assets are operated and exploited to attract international attention to twenty four major international open data collections. These collections will exploit Australia's research data infrastructure and Australian research institutional data capabilities to enable the best possible international profile and partnerships.

These collections are intended to be highlighted in a Festival of Open Data to be held in June 2015. This should showcase many of Australia's research data assets, and their value in engaging major Australian objectives and international partnerships.

Open Data Collections (ODC)

ANDS will also support the smaller, medium sized and regional universities that had not received the MODC offer. ANDS offered 14 universities $35,000 each to make a significant data collection openly

available and well described through their institution. The funding has been provided through the University libraries in order to enable the library to take on a substantial role in research data, in collaboration with the other stakeholders at the universities.

Eleven institutions responded accepted this offer, and have agreed on the contracts. These are:

- Australian Catholic University
- Bond University
- Charles Sturt University
- Edith Cowan University
- Federation University
- Southern Cross University

- University of Canberra
- University of New England
- University of Southern Queensland
- University of the Sunshine Coast
- Victoria University

These universities have not yet identified which data collection they will make available using the funding. This will take place through a process in which ANDS will be connecting with the university to help them in identifying the data collection and the actions that need to be taken in the course of the project.

Community Engagement

As well as supporting institutions with funding and expertise through its main funded projects, Major Open Data Collections (MODC) and Open Data Collections (ODC), a number of tailored workshops were initiated by ANDS during the January to June 2014 period. The University of Western Australia, Curtin University, Edith Cowan University, Charles Darwin University, James Cook University, and University to Tasmania were contacted to determine if they required support from ANDS with training, advocacy, or awareness building in managing research data. Other institutions will be approached during in the second half of 2014.

A series of visits to the smaller and regional universities in Southern and Central Queensland have taken place, which has resulted in the need for an intensive training session for staff members at these and other smaller and medium sized universities. The Institutional Engagement team has continued one on one engagements with universities through a series of targeted visits and conversations to determine their specific needs, see section 4.2 for details.

4.1 Highlights, Difficulties and Breakthroughs

Between January and June 2014, of the 24 institutions approached to establish a Major Open Data Collection, project charters and contracts were finalised, or in the process of being finalised, with the following 15 institutions;

- Monash University (MOU)
- La Trobe University
- University of Melbourne
- RMIT University
- Swinburne University
- Flinders University

- University of Adelaide
- University of South Australia
- University of Western Sydney
- Queensland University of Technology
- Griffith University
- University of Queensland

- University of Western Australia
- Curtin University

Three institutions; University of Tasmania and James Cook University, University of Technology Sydney have selected their Major Open Data Collection and finalised the project charter, and the contract is being reviewed by their institution's legal area.

At this stage Deakin University is the only institution that has not taken up, or formally declined the offer. It is anticipated that a project to establish a MODC at Deakin may not go ahead.

Institutions have selected a diversity of collections for the MODC engagement and include the following which have both national and international significance;

	Institution/State	MODC Project Summary
1	*Monash (VIC)*	This project will deliver two series of significant collections related to cancer and diabetes, along with other diseases. These collections will be sourced from; 1) Significant imaging instruments hosted or owned by Monash. These will include selected instruments from the Monash Medical Imaging suite, the Australian Synchrotron and other centres or facilities where possible, as part of a larger Imaging Locus initiative at Monash and 2) From accessible discipline specific repositories using tools funded under previous ANDS projects Our vision is to deliver a large number of public datasets in areas of particular interest and value to the University (such as the newly formed Sub-Faculty of Biomedical and Psychological Sciences and the Monash Micro Imaging (MMI) facility). Data will be captured from at least 12 Monash electron microscopes during this project, with instruments to be included if they come online during the period of this ANDS project. There is a plan to have several additional new instruments commissioned in August/September. To achieve these goals, the project will comprise three interrelated areas of work; 1) Work with the creators and owners of the data to make existing and newly created datasets freely available online, either directly or via a click through registration. 2) Work with the owners of a range of imaging instruments and platforms currently located at Monash to implement policies that will ensure that data produced by these instruments is made freely available with a proscribed time. 3) Create a sustainable operational environment, including the maintenance of software funded in previous ANDS projects to enable improved data and metadata capture, storage of metadata

		and data and enhanced discoverability of the data and metadata, all of which will improve access to the data collections. This work will be undertaken by a collaboration of staff from across the University, including eSolutions, Library, Monash eResearch Centre (MeRC) and the recently established Bioinformatics Platform.
2	*La Trobe (VIC)*	The project aims to build a Healthy Communities Research Data Collection. This collection will consist of data sourced from a number of local health and welfare providers in Victoria, with which La Trobe University has been collaborating. These data will be brought together to produce a major open access data collection consisting of epidemiological data, general health and wellbeing data, health services utilisation data. This project will focus on compiling an open access population health database for the City of Greater Bendigo, bringing together health and wellbeing data from the Medicare Local in Bendigo and Bendigo Health Service, coupled with relevant available data from the ABS, AIHW and the Municipal Health Plans. Work will be done to ensure only non-identifiable data is added to the collection and to link related information through linked data and open data principles.
3	*University of Melbourne (VIC)*	This project will use one nationally significant life science data collection (*Drosophila*) as a demonstrator for the verification, publication, preservation and reuse of the research collection as an open collection; and as a use case for long-term access through preservation planning.
4	*RMIT (VIC)*	This project works with research data from MicroNano Research Facility and Health Innovation Research Institute. Project aims include to: 1) consolidate existing disparate collections 2) demonstrate integrated processes and quality assessment procedures 3) curate workflows and experiments as generic research data Overall the aim is to expose collection information in a form that can be rapidly searched, found, accessed and utilised by external and internal researchers.
5	*Swinburne (VIC)*	This project would see the implementation of Digital Object Identifiers (DOIs) to the Policy Online data set in order to make valuable scholarly material more discoverable, citable, and usable.
6	*Flinders University (SA)*	This project would make existing data collected by Flinders palaeontology researchers available for open access. The project will initially deal with specimen image, dimension, location and geochemical data. This project and ANDS funding will be used to: 1) Standardise metadata schema across existing datasets 2) Create and add metadata to associated images and link these to

		other compiled data
		3) Negotiate permissions to use other related datasets
		4) Aggregate data to allow complex interrogation
7	*University of Adelaide (SA)*	As genomics and microbiome data are becoming routinely generated for ancient samples, it becomes crucial to keep track of raw data, processed data, bioinformatics pipelines and metadata in an organised and searchable fashion. This project will build a world-first structured database for ancient human DNA, integrating all layers of available genomic and microbiome data and associated metadata in a searchable format. The database will provide open access to datasets derived from ancient human samples studied at the Australian Centre for Ancient DNA (ACAD).
8	*University of South Australia (SA)*	The FLOW project will undertake research to develop a metadata framework for the collection, description and cataloguing of water quality data. In parallel with the research the project will develop open source software which applies the metadata framework and collects water quality data. Whilst the project is focused on water quality data the metadata framework will not preclude extension in the future to include other forms of water data. The FLOW project will bring together a diverse group of researchers and information services professionals, including SA Water.
9	*University of Western Sydney (NSW)*	This project is designed to assist in the creation of a major research data set supporting "The Project for Western Sydney", in the context of ongoing work at the University of Sydney (UWS) to create a Humanities Data service associated with the Digital Humanities Research Centre. This project will focus on data-discovery and cataloguing of Digital Humanities and Social Science data that will assist in the delivery of the Project for Western Sydney.
10	*Queensland University of Technology (QLD)*	Deliver a set of spatially integrated datasets that demonstrates and enables researchers and government to answer research questions within multiple jurisdictions. Data will be augmented from international partners, derived data, and data generated by the Queensland University of Technology (QUT). This data will be well described, connected, published and made citable. The project is a sub-project to the larger Queensland Globe project in which the QUT, specifically the Institute for Future Environments, is commissioned to provide a Queensland Government G20 Globe for the G20 Summit event.
11	*Griffith University (QLD)*	This project will develop a significant collection of data and information integrating and aggregating data with a focus on place. It will deliver an online collection of data and information specifically designed around a regional council authority boundary. A small number of regional authorities will be used in the initial rollout using two focal points; the first being a community climate change adaptation response, and the second promoting

		healthy communities.
12	*University of Queensland (QLD)*	Undertake two projects that deliver major and significant collections of data. These projects are:
		Australian Wheat - to establish a T3 database for Australian wheat breeding data and host this database at the Australian node of the wheatIS based at UQ.
		Flint Collection - to develop detailed annotations of the recordings and to link the annotations to the recordings; and to connect the recordings and their annotated transcriptions to a user-friendly platform for public access.
13	*University of Western Australia (UWA)*	Due to the tight relationship between subcellular compartments and protein function, proteome-wide protein localization is an important goal in cell biology for plants and crops. For one model plant, *Arabidopsis*, a wealth of computational and experimental data exists that has been used to create global protein localisation data, and protein relationships within the SUBA database that has been built and maintained at UWA. Increasing demands for large-scale location and relation data for systems biology in both model plants and agricultural plant species indicate that expansion of data capture, integration and reuse for localisation analysis is now required. This project aims to build a novel protein localisation database environment and web interface for plant and crop scientists. The data portal will allow users to access captured genomic, proteomic, biochemical data as well as newly generated global localisation data, localisation interpretation and relation data across 5 crucial agricultural crops in addition to Arabidopsis. The data access and visualisation will significantly increase the use and reuse of these data across an international audience of plant and crop scientists and systems biologists.
14	*Curtin University (WA)*	The John de Laeter Centre for Isotope Research (JDLC) is a Perth-based multi-institutional research infrastructure centre providing the academic, resources and environmental research sector with advanced analytical facilities and expertise. During 2014, the Centre will commission a number of new instruments generating data relating to digital mineralogy and materials characterisation for pure and applied research. The new mineral analyser instrument will be used on up to 2,000 samples from the Geological Survey of Western Australia (GSWA), characterising the mineralogy of the samples into searchable digital mineralogical and geochemical datasets. These datasets will be made accessible.
		This ANDS project will create an appropriate metadata schema for these datasets, capture and enhance the metadata for a 150 sample subset of the collection under that schema and make that metadata available to Research Data Australia and the AuScope portal to facilitate discovery and

		access to the datasets by the international research community.
15	*University of Technology, Sydney (NSW)*	Establishment of Dharmae – a Data Hub of Australian Research on Marine and Aquatic Ecocultures

Table 1. MODC Project Descriptions

Delays in compiling and executing contracts have been due to the institutions not being able to select the Collection to focus on. This has been a difficult decision for some institutions as they have strengths in multiple disciplines. The result of these delays means institutions will have less time in which to complete their projects in order to deliver a Significant Collection by the end of May 2015.

4.2 Activities Undertaken

ANDS staff and the institutions have continued to get MODC projects finalised and contracted. As well, ANDS staff have continued to offer expertise into the projects upon request.

Tailored workshops were initiated by ANDS during the January to June 2014 period. A number of institutions were contacted to determine if they required support from ANDS with training, advocacy, or awareness building in managing research data at their institution. In response to this offer workshops were developed in collaboration with the Capabilities teams, who ran the workshops on-site at the university, with input from the Institution. Workshops were held in June at University of Western Australia, Curtin University, Edith Cowan University, and James Cook University and the Australian Institute of Marine Science (AIMS). A workshop at the University of Tasmania is planned for the end of July.

ANDS has continued engagements with universities through several means including:

- The NSW roundtables held every 6-8 weeks in Sydney in collaboration with Intersect
- The Victoria/Tasmania Informal's held every 6-8 weeks at one of the institutions in those states
- The Research Data Management online catch-up sessions, targeted at the smaller, medium sized and regional universities. Discussions at these sessions are focused on the challenges specific to these smaller universities

4.3 Performance against agreed Key Performance Indicators

KPI	Proposed Activity	Progress
1	The number and coverage of data repositories providing metadata feeds to the national registry compared to the number of data repositories.	Target: ANDS intends to maintain at least 100 metadata feeds. This will cover at least 60 research data-holding institutions. Result: 71 institutions fed collection descriptions to RDA along with 46 individual collections. From these institutions, 157 data

		source feeds have been setup (101 automatic and 56 manual feeds)
2	The number and coverage of institutions and number of research groups with which ANDS has engaged.	Target: ANDS will maintain engagement with all Australian universities, PFROs, and 20 major Government data providers this year, and through them at least 50 research groups. Engagement with 38 institutions via its Major Open Data Collections (MODC) or Open Data Collections (ODC) funding program. Result: Developing and providing tailored workshops for six institutions. Developing intensive workshops for 14 small and medium size institutions Responding to enquiries from our institutional partners on an ad-hoc and providing expertise on an as needs basis.
3	The number of institutions with research data management policies and practices consistent with ANDS recommendations: 31	Target: 31 research institutions will maintain research data management policies and practices consistent with ANDS. Result: 20 – Australian Catholic University, Australian National University, Bond University, Charles Darwin University, Deakin University, Edith Cowan University, Griffith University, La Trobe University, Monash University, Queensland University of Technology, Southern Cross University, The University of Melbourne, University of Canberra, University of New England, University of Newcastle, University of Queensland, University of Tasmania (IMAS), University of Technology Sydney, University of Wollongong, Victoria University. ANDS has engaged with many more institutions on the development of research data management policies and practices this year, however comparatively few of those engagements have been finalized at this stage.
4	The satisfaction of researchers and partners (see below) with ANDS services as measured by an annual survey.	The annual surveys for researchers and partners have been prepared. They will be sent out in July and the report based on the responses will be drafted in August.

5	The number of data access and sharing agreements with stakeholders – principally research institutions, government data agencies, government research agencies.	Target: ANDS will maintain at least 50 agreements to make information available. Result: 71, with 3 new agreements added since June last year.
6	The number of research data sets in the ARDC.	Target: more than 60,000 collections Result: June 2014: 97,672 collections are made available in Research Data Australia as at 30 June 2014. 89% of the total collection records came from the top 5 top contributors: 22% from Geoscience Australia (GA), 19% from Queensland Facility for Advanced Bioinformatics (QFAB), 17% from State Records NSW, 16% from Public Records Office Victoria (PROV) and 15% from Australian Ocean Data Network (AODN). The remaining 11% of the total collections are from 66 contributors with a total of 85 corresponding groups appearing in Research Data Australia. The number of ANDS research data sets increased 45% from June last year.
7	The number of research data sets with persistent identifiers.	Target: more then 14,000 Result: 12,082 persistent identifiers (9,677 PIDS handles, 2,405 DOI handles), 17% higher than June last year. These identifiers were minted by 17 Identify My Data (PIDS) clients and 13 out of 29 registered Cite My Data (DOI) clients.
8	Assist in delivering, by June 2015, research data collections which are open and accessible and significant to research institutions, the nation, and the international research community.	Fifteen *ANDS NCRIS 2013 Subcontract Agreement* pertaining to Major Open Data Collections (MODC) projects between ANDS and the institutions have been finalised or in the process of being finalised; awaiting final signatures– see section 4.1 for details. Eleven *Monash University Contractor Agreements* pertaining to Open Data Collections (ODC) projects have been finalised – see section 4.1 for details.

5 International Collaboration Program

The scope of this program includes collaboration with data infrastructure providers to ensure that Australian research data infrastructure is compatible with international approaches.

Until 2012, international collaboration activities within ANDS occurred across a number of different programs, including for example, selection of DataCite as our preferred Digital Object Identifier solution, and engaging with international partners to adopt our technology as a way of future proofing the investment and expanding the pool of potential developers.

The Research Data Alliance has now become the main focus for ANDS' involvement in international activity; although a number of existing activities will continue in parallel. The purpose of the Research Data Alliance is to accelerate international data-driven innovation and discovery by facilitating research data sharing and exchange, use and re-use, standards harmonization, and discoverability. This will be achieved through the development and adoption of infrastructure, policy, practice, standards, and other deliverables. The Research Data Alliance has been founded through commitments by Australian Commonwealth Government, the European Commission, and United States of America. The work of the Research Data Alliance will primarily be undertaken through its working groups and interest groups. Participation in working groups and interest groups, starting new working groups, and attendance at the twice-yearly plenary meetings is open to all.

In this period, Australia will fund the position of the Research Data Alliance Secretary General. This is both important to the Research Data Alliance in its first period of operation, but also particularly important in Australia's relationship with the US and Europe, cementing the leading role of Australia in research data internationally.

Note that this program will be conducted using a combination of NCRIS, EIF, CRIS, and NCRIS 2013 funds, as well as separate funding under a contract entitled "Data Web Forum".

5.1 Highlights, Difficulties and Breakthroughs

The major highlights involve the Research Data Alliance:

- Australia co-hosted the 3rd RDA Plenary
- The secretary General has been recruited
- RDA, and Australians involvement continues to grow

5.2 Activities Undertaken

As planned, ANDS has contracted to fund the Research Data Alliance Secretary General. After an international search, Mark Parsons was selected by the Research Data Alliance, with Ross Wilkinson as chair of the selection committee. As a result, a contract was written with Rensselaer Polytechnic Institute to employ Mark Parsons, and pay associated travel and operational costs for 1 year. ANDS staff have also been active in the RDA Council, Technical Advisory Board, Secretariat, and Working Groups.

Most notably, ANDS co-hosted the 3rd RDA Plenary held in Dublin in March 2013, though this activity was not funded using NCRIS 2013 funding, and will thus be reported separately.

5.3 Performance against agreed Key Performance Indicators

[To be updated]

6 A summary table of actual expenditure for the period covered by this report

Program Title	Actual Expenditure for the period
National Collections	
operation, management and governance costs	$ 13,277
salaries and on costs for technical staff	$ 6
infrastructure maintenance	$ -
utilities	$ -
rent	$ -
consumables	$ -
other	$ -
National Services	
operation, management and governance costs	$ 33,249
salaries and on costs for technical staff	$ -
infrastructure maintenance	$ 44,250
utilities	$ -
rent	$ -
consumables	$ -
other	$ -
Institutional Engagement	
operation, management and governance costs	$ 74,431
salaries and on costs for technical staff	$ 7,652
infrastructure maintenance	$ 674,600
utilities	$ -
rent	$ -
consumables	$ -
other	$ -
International Collaboration	
operation, management and governance costs	$ 7,386
salaries and on costs for technical staff	$ -
infrastructure maintenance	$ 321,198
utilities	$ -
rent	$ -
consumables	$ -
other	$ -
TOTAL EXPENDITURE	$ 1,176,049
INCOME RECEIVED	$ 2,886,993
BALANCE REMAINING	$ 1,710,944