ANDS Project Final Report

for

ReDBox

Prepared by Duncan Dickinson

Document Version 0.1

ANDS Project Code	EIF048
Organisation responsible for the project (Subcontractor)	*QCIF*
Organisation that has undertaken the work (Sub-Subcontractor)	*Queensland Cyber Infrastructure Foundation(QCIF)* *(for provision of project resources)*
Name of Contact Person	*Duncan Dickinson*
Project Duration Period	*18/06/2011 - 18/06/2012*
Author(s) of this Report	*Duncan Dickinson*
Role(s) of Report Author(s) within the Project	*Project Manager*
Reporting Period	*18/06/2011 - 30/06/2012*

PART A – FINAL PROGRESS REPORT

This section of the document provides a framework for the subcontractor to describe the progress in this final reporting period towards achieving the goals of an ANDS funded project.

1. Summary of project progress during this reporting period

- *Please provide a short overview (1-2 paragraphs) of project progress during this reporting period.*

- *Where relevant, please include any changes in staffing, your institutional context and/or any progress or issues with project partners; as well as release of documents, software code and communication/dissemination activities with stakeholders and/or the wider community in this reporting period.*

- *Where relevant please include information about your progress or issues regarding your sustainability plan.*

Since the last progress report (Feb 2012), the project has undertaken the following work:

- Released ReDBox version 1.5:
    - Included two rounds of user acceptance testing (UAT) with the community
    - For functionality details, please refer to http://code.google.com/p/redbox-mint/wiki/ReleaseHistory
- Organised and led the ReDBox Community Day
    - This was very well received and generated a number of feature requests: http://www.redboxresearchdata.com.au/feature-planning
- Worked with a number of universities regarding their ANDS Metadata Stores project planning

The UAT work for the version 1.5 release was led by Vicki Picasso and included a number of universities. A number of issues were raised through the release process and these were managed via our issue tracking software.

Current (known) implementation sites include:

- University of Newcastle
- Flinders University
- Central Queensland University

ReDBox is currently being explored by a number of institutions, as indicated in Simon Pockley's posting to the ANDS Partners list: https://docs.google.com/spreadsheet/ccc?key=0AoBX4iCLq9QHdEIxWUFOU3ZOcXlaTXNNdDQzY19GVnc#gid=0

A user list has been setup at http://www.redboxresearchdata.com.au/users

2. Milestone Reporting

Please complete the table below to outline work undertaken toward achieving each agreed project deliverable.

Deliverable [§]	**Summary of work undertaken in this reporting period towards achieving this deliverable.**	**Documentation** [§] *Where appropriate, please provide documentation that illustrates progress using a suitable method e.g. by reference to an attached document, a URL pointing to an on-line document, a URL and accompanying access information to a test/deployed software system etc)*	**Status*** *Please indicate using an 'X'* **NB. It is expected in a Final Report that all deliverables will have a status of 'Complete'**				
			On schedule	Slippage	Delayed/Impact	Complete	Yet to Start
D1	D1: Project Plan	https://jira.ands.org.au/secure/attachment/14701/EIF048_QCIF_PMP_v1.0_20111012_NN_ND_MFW.docx				x	
D5	D5: Code enhancements	Version 1.1 -> 1.5 released under EIF048. See http://code.google.com/p/redbox-mint/wiki/ReleaseHistory				X	
D7	D7: Updated documentation Resulting updated source code deposited in open source repository	Code and tech wiki: http://code.google.com/p/redbox-mint/ Public doco: http://www.redboxresearchdata.com.au/documentation				x	
D9	D9: First Progress report	https://jira.ands.org.au/secure/attachment/14503/ands-progress-report-EIF048.docx				x	
D9	D9: Second Progress Report	https://jira.ands.org.au/secure/attachment/17526/EIF048_ReDBox_Progress-report-EIF048_v0.2.docx				x	
D10	D10: Acceptable final report						x
D11	D11: Deployments of code					x	
etc							

3. Collaboration and Support

- *Are there any issues that you would like to discuss with other ANDS-funded projects where ANDS could act as a facilitator?*
- *Are there any issues where you would like direct advice and support from ANDS?*
- *Are there any achievements during this reporting period you would like ANDS to share on your behalf?*

EIF048 has been highly collaborative throughout the project. The project board included representatives from the project, ANDS, University of Newcastle and Flinders University. Details of the project were posted to the ReDBox website: http://www.redboxresearchdata.com.au/governance/eif048

The ReDBox community is now establishing user and technical groups so as to organize further work and ensure the ongoing utility of the code base.

PART B – OVERALL PROJECT SUMMARY

This section of the document provides a framework for the subcontractor to reflect upon all work undertaken during this project, highlighting successes and challenges that have occurred during the entire lifespan of the ANDS funded project, both from a general/lay and expert perspective.
NOTE: The content of PARTS 4 or 5 may be used to inform subsequent ANDS public communications such as press releases, website content etc.

4. Lay Summary

ReDBox (Research Data Box) provides institutions will a research data catalogue that helps them meet federal and institutional requirements. Assisted by a name authority system (Mint), ReDBox is an extensible system for use by libraries and research office to integrate with internal research data management processes and national programs. For further details, please refer to http://www.redboxresearchdata.com.au/documentation/system-overview

For institutions undertaking an ANDS Metadata Stores (AMS) project, ReDBox provides a solid technical foundation for meeting the mandatory requirements and many of the optional items. The table below provides an overview of where ReDBox meets the AMS deliverables and where further work may be required.

Deliverable	Mandatory	ReDBox Functionality (Version 1.5)
D1: A working feed of records describing collections and associated activities, parties and services to Research Data Australia, in the current version of RIF-CS (1.3), demonstrated to meet the quality requirements for RIF-CS records set by ANDS. (realises F3) (realises F6)	Y	• As of version 1.4: o ReDBox provides RIF-CS 1.3 feed for collections o Mint provides RIF-CS 1.3 feeds for activities & services • Mint provides EAC-CPF feeds (NLA) for parties o This is in addition to RIF-CS Party feeds
D2: Coverage in the Research Data Australia records feed of collections from at least three distinct faculties (or equivalent organisational units) within the institution (realises F1)	Y	• This will be a site-based decision
D3: Demonstrated alignment of metadata records about parties with an institutional name authority (HR or Library), with the authoritative form of the name sourced external to the metadata store, and with new researcher descriptions added to the metadata through regular update from the	Y	• Mint provides an external name authority service • Mint can manage multiple identifiers (internal and external) for each party • Mint can currently accept and harvest batch updates via a CSV upload and this can be automated as needed

Deliverable	Mandatory	ReDBox Functionality (Version 1.5)
name authority (realises F7)		• Mint's various integration points allow for systems other than ReDBox to interact with the name authority • Sites will need to determine internal data sources for identities
D4: Demonstrated alignment of metadata records about parties with the ARDC Party Infrastructure Project, with researcher descriptions contributed to the NLA, and with People Australia identifiers for researchers recorded against researchers (realises F8)	Y	• Integration with the NLA ARDC-PIP was release in version 1.5
D5: Demonstrated alignment of metadata records about activities with institutional and external sources of truth (Research Office, ARC and NHMRC grant registries), with the authoritative description of the activity sourced external to the metadata store, and with new researcher project added to the metadata through regular update from the sources of truth (realises F7, F8)	Y	• Mint can manage internally identified activities and is able to provide information from external sources of truth. • Activity information can be uploaded and harvested into Mint via CSV files from pertinent data sources within the institution
D6: Demonstrated process for registering new collections in the university; this can include automated update, or semi-automated (notification-based) (realises F9)	Y	• ReDBox provides systems for the manual description of research data through a workflow and form-based interface. • ReDBox also provides an "alerts" system that allows ReDBox to become aware of collections created by other tools. • ReDBox alerts can also be created by research management systems to signal that research projects may need data curation (for example, towards the end of a grant)
D7: Demonstrated ability to manage the following aspects of the collection lifecycle through recording and exposing relevant metadata: (realises F10) • embargo dates for collections, where applicable • current online location of collection (on internal store or external store) • current offline location of collection • intellectual property rights (licensing, restrictions on reuse)	N	• Embargo functionality is not currently available in ReDBox but is on the feature plan (link) • Online and physical locations of data can be recorded in ReDBox • Rights information is captured in ReDBox • Retention and deposit information is captured in ReDBox • Support files can be uploaded with ReDBox metadata, allowing for the storage of associated data management and ethics documents

Deliverable	Mandatory	ReDBox Functionality (Version 1.5)
• retention policy (disposal date, deposit date) • policy framework (data management plan relevant, ethics forms relevant)		
D8: A public researcher or research profile portal, exposing publishable metadata about the research data being held at the institution (realises F2)	N	• Institutional customisation needed • ReDBox has a customisable user interface for searching across the collection metadata. • Mint can be customised to provide a party, service and/or activity profile system, disseminating associated collections. • Mint has a number of integration points that allow an institution to collate information from a number of sources. o e.g. Mint can be used as a name authority in the IR and the IR, can supply Mint with citation information for researchers. • Mint and ReDBox can both provide feeds to other discovery and profile systems (such as Vital and Vitro)
D9: Demonstrated ability to feed a selected subset of the collection records relating to a particular discipline to a discipline registry, following the metadata schema and conventions of that registry (realises F4)	N	• ReDBox is able to meet this deliverable. Metadata elements can be used to determine the subset of collections required and metadata exchange with discipline registries can utilise specialised formats and schemas.
D10: Demonstrated ability to manage the following aspects of the collection lifecycle through • recording and exposing relevant metadata: (realises F10) • citation requirements (authoritative identifiers, including DOI, preferred citation format) • citation tracking of collections • audit information (refer to publications audit) • proprietary tools and formats used in collecting the collection	N	• ReDBox and Mint provide a templating system for disseminating metadata that can easily be customised and new formats/schema provided • ReDBox provides fields for citation entry • DOI implementation was held out of version 1.5 due to ANDS delays • ReDBox and Mint were built from an understanding of linked data and aims to use globally unique identifiers wherever possible o Handles can be used to identify Mint and ReDBox objects o DOI functionality can be developed now that ANDS has released the service • Citation tracking has not been discussed for ReDBox
D11: Strategic reporting on contents and	N	• Statistics functionality in ReDBox and Mint will

Deliverable	Mandatory	ReDBox Functionality (Version 1.5)
coverage of metadata store for internal use (realises F11)		need to be scoped (link)
D12: Storage and exposure for discovery of object level metadata, and alignment of object level metadata with collection metadata (i.e. ability to navigate from object metadata to collection metadata; update of object metadata aligned with update of collection metadata) (realises F5)	N	• ReDBox provides an XML ingest system, allowing other systems to post collection metadata to ReDBox. • ReDBox's storage layer can handle most file formats • This is typically provided once data has been published • Functionality for the updating of ReDBox metadata from other system would need to be scoped.
D13: Storage and management of technical metadata for object and collection reuse, including software and equipment descriptions, methodology, and data interpretation (realises F5)	N	• As above • This is most likely out of scope for ReDBox. It is suggested that this type of metadata be stored and managed by displine/project-level systems

Project public information

- Project website URL (if any): http://www.redboxresearchdata.com.au/
- Primary contact information: Duncan Dickinson (d.dickinson@qcif.edu.au)
- EIF048 Team members: Duncan Dickinson, Greg Pendlebury, Norm Lawler
- Collaborators: Vicki Picasso (Newcastle) & Amanda Nixon (Flinders)
- What documents, information, guides, policies or software did this ANDS project deliver, and where can they be accessed?
    - Public website: http://www.redboxresearchdata.com.au/
    - Technical website: http://code.google.com/p/redbox-mint/
    - Project tracking site: https://www.pivotaltracker.com/projects/316187
- How can people publicly access your shared data?
    - Public demonstration site: http://demo.redboxresearchdata.com.au

5. Expert / Technical Summary

This information is available via a number of resources:

- ReDBox documentation: http://www.redboxresearchdata.com.au/documentation
- ReDBox technical documentation: http://code.google.com/p/redbox-mint/
- ReDBox mailing list: http://groups.google.com/group/redbox-repo/
- ReDBox technical mailing list: https://groups.google.com/d/forum/redbox-dev

6. Lessons Learnt

- *Describe your top 3 lessons learnt from this project.*
    1. The area of research data management is very much in a state of flux. It's difficult to pin down processes and developments (e.g. ORCID v NLA ID, the role of DOIs, data sharing metrics). In the Cynefin model[1], much of what we're doing is in the Complex domain and requires exploration and even readiness to try a solution for a period but drop it if it fails (a try early, fail early approach).
        1. Work and discussions in the community demonstrate that people are identifying areas in which a move to the Complicated domain will occur as patterns settle.
    2. Linking into a global approach to the identification and sharing of data is still emerging and this pushes software and processes to think beyond a local context
        1. Are we at a point where the UK/Europe, US, Australian approaches can be compared and commonalities be found?
    3. The work we're all doing in this arena really has a mid-term (3-5 years from now) payoff as it includes culture, policy and network changes. This means that the value of the work may not be obvious to stakeholders in the near future. This means we have time to get the models close to "right" but collaboration is essential to ensuring that the best ideas/approaches are taken on-board.

- *Describe your top 3 issues and how you would recommend planning to eliminate these issues in future activities.*
    1. Lack of cross-project planning: there are a number of intersecting ANDS projects that do not appear to have been aligned to ensure that work can be planned and carried out effectively. In the EIF048 example, better alignment across the following programs would have been extremely useful:
        1. Metadata Stores
        2. NLA ARDC PIP
        3. Vocabulary services

[1] http://en.wikipedia.org/wiki/Cynefin

4. DOI M2M provisioning

How to resolve this is a complex issue with so many projects and stakeholders. It is likely that, under the Metadata Stores projects, these will start to "gel" more.

2. A question surrounds metadata store sustainability across Australia. Whilst the ReDBox community is largely transparent and communicative, the same cannot be said about all other metadata stores. The goal should be to align all efforts where sensible and, to date, this has not happened at any basic level. This, however, is beyond the role of the ReDBox community. A cross-community service would assist here and ANDS may be well-placed to co-ordinate this.

I would also like to take an additional and slightly different angle here and list 3 sustainability challenges for the ReDBox community post-EIF048:

1. User group: The formation of a ReDBox user group to determine the scope and development of the various features under consideration for the metadata stores activity
2. Technical panel: To create a governance model for the ongoing development of the ReDBox codebase.
3. Post-metadata stores sustainability: we saw the institutional repository community fade once the ASHER money was no longer available[2]. When ERA came in, many universities undertook work on their IR but the level of communication was much lower (as ERA deadlines loomed closer).
    - The closed nature of CAIRSS also impacts on the ability of those outside the library sector to contribute. An open, combined data + IR community is likely to be needed.

7. Sustainability Plan

Three efforts will assist ReDBox in terms of sustainability:

- QCIF will service paid maintenance contracts, providing a core technical staff
- QCIF (and potentially other organisations) will also provide consulting services, aiding organisations that wish to outsource feature development
- The Metadata Stores community subset that selects ReDBox will start engaging in feature development and customization. Work is being undertaken to ensure that this happens in the most effective manner possible.
    - This will be aided by the user and technical groups described in Section 6.

The ReDBox code base has always been open source and provided through a public repository: http://code.google.com/p/redbox-mint/source/checkout . The Java/Maven modules are posted to the public repository: http://search.maven.org/#browse%7C808788309

[2] Alternatively, the software met all needs and stabilised at that point

8. Availability of Project Material

Please refer to Clause 14 ("Project Material") of Schedule C (i.e. the EIF or NCRIS Funding Agreement between DIISR and ANDS/Monash University) of your contract, to which all subcontractors to ANDS/Monash University are bound. As outlined, it is expected that project materials developed with funding from this agreement be made available for non commercial use.

ANDS recommends licensing of your project materials using an AusGOAL (Australian Governments Open Access and Licensing Framework) licence, which makes it clear to others the terms of re-use within the boundaries of copyright law.

- *Please provide a brief statement of project materials produced (e.g. documentation, guides, policies, software) and how that material has been made available for re-use (including the terms for re-use).*

All code is covered by the GPL (v2) and documentation is licensed under Creative Commons. Links to these resources have been made throughout this report but http://www.redboxresearchdata.com.au/ is the primary access point.