CAUL Review of Australian Repository Infrastructure



CAUL Fair, affordable and open access to knowledge program

September 2018



Project Steering Committee Members

	Ginny Barbour	Director, AOASG	Julia Hickie	Assistant Director, Trove Data, Discoverv and Deliverv. NLA
		(Chair)	Harry Rolf	CAUL Communications Officer
	Katrina Dewis	Associate Director (Acting), UTas		(Support)
	Janet Fletcher	University Librarian, Victoria University, Wellington	Alexander Sussman	Associate Director, Academic
M	Maude Frances	Associate Director, Library Digital Repositories, UNSW		Services, UNSW
			Natasha Simons	Program Lead, Skills Policy and
	Andrew Harrison	Research Repository Librarian,		Resources, ANDS
		wonash oniversity	Belinda Tiffen	Director, Library Resources Un UTS

ources Unit,



Project Group Members

Mary O'Connor	eResearch Support Project Librarian, Adelaide	Margaret Pembroke
Kay Steel	Manager, Research and Strategic Projects, Federation	Paula Callan
Scott Abbott	Manager, eResearch, eScholarship, UTS	Katya Henry Elke Dawson
Eleanor Colla	Researcher Services Librarian, UNE	Allison Brown

Digital Scholarship Librarian (Systems) SCU Scholarly Communications Librarian, QUT Research Support Librarian, QUT Manager, Open Research, ANU Senior Research Publications and Data Librarian, UNE



Project Products

- Review and report on the current Australasian research repository infrastructure.
- Review and report on the international research repository infrastructure and developments.
- 3. Develop a series of repository user stories.

1.

6.

- Recommend and make improvements to the current Australasian research repository infrastructure (improve and make the most of what we have).
- 5. Develop and report on an ideal state for Australian research repository infrastructure.
 - Investigate and make recommendations for next generation repository tools (consortium approach, possible infrastructure project).
 - Investigate and make recommendations for a possible
 "Research Australia" collection of research outputs.

Observations

#1 Review and Report on the current Australasian institutional research repository infrastructure

Online survey

No overlap with the CAUL 2017 Repository Survey

45 / 48 institutions responded

Observed under reporting

Repository infrastructure and content

Question: Briefly describe your current repository infrastructure and the content it holds



ASHER funds 2007 Aged solutions Commercial alignment figshare 4, maybe 7?

Figure 1: Percentage of institutions using repository software.

Content	# Institutions / 45
Research outputs	45
Theses	31+
Research data	22+
Archival and library collections	16
Journals	5
Images and multimedia	3
Course materials, readings	8

Open Access

Questions: Does your institution have a policy that mandates Open Access?

Open Access Policy	# Institutions / 45
OA policy	16 (36%)
OA statement / partial	15 (33%)
No OA policy	14 (41%)

Grant funder OA policies - support for and monitoring

Question:

Grant funder OA policy

- a): Does your repository support OA policies of funding bodies?
- b): Is policy compliance monitored by whom/how?

Suports	# Institutions / 45
Grant funder policies	33 (66%) rising to 43 (95%)
Monitor compliance	14 (31%) rising to 19 (42%)

Research Activity Identifiers

Question: Does your repository use Research Activity Identifiers.

Uses RAI	#Institutions / 45
Yes	5 (11%)

Repository Interoperability and Integration

Question:

Is your content harvested by any of these services?



Discoverability with search engines and portals

Repatriating Discipline specific repository records

Are you collecting or planning to collect records from discipline repositories for your institutional repository? For example, arXiv.org, rePEc: Research Papers in Economics,

Collecting ?	# Institutions / 45
Yes	25 (56%)

Preservation strategy

Question:

Do you have a preservation strategy for the repository content?

Preservation strategy	# Institutions / 45
Yes	13 (29%)

Review and report on the international research repository infrastructure and developments. WG # 2

Methodology

- Brainstorming on known examples of good practice among the group. Initial source of examples was derived from work in 2017 done by the previous AOASG and CAUL repository working group
- Examples included were refined through an iterative process through discussion within the group
- Detailed information on relevant initiatives and specific practices was compiled from relevant websites, including COAR, ROARMAP, personal contacts at the initiatives
- Classified examples into three broad categories of "Infrastructure", "Tools" and "Supporting Organisations".

Findings

Infrastructure organisations

- La Referencia
- OpenAIRE
- OpenDOAR
- PubMed Central
- SHARE

Tools

- CORE
- Dataverse
- Dryad
- Figshare
- IRUS-UK
- Re3data
- ResourceSync
- RIOXX
- Scholix

Supporting organisations

- CARL
- COAR
- LERU
- LIBER
- National INstitute INformation
- SPARC
- UK-CORR

Key themes

- Most important concept is interoperability
- High degree of co-dependence
- Most successful initiatives had stable ongoing funding
- Local and regional networks are key

Caveat

This is a constantly changing landscape and this review is just a snapshot in time.

There is a need for regular scanning of the environment to keep abreast of developments.

Develop repository user stories

WG # 3



Stakeholder groups

- What can a repository do (for me)?
 - Shifting capabilities, requirements and stakeholders
- Who are the stakeholders?
 - Librarians Researchers Funders -Administrators - Government - NGOs -Publics - Teachers – Students - Systems
- User stories
 - Matrix captures diverse uses existing and aspirational

As a depositor, I want the deposit process to push my work to ORCID to remove duplication of effort

 As a librarian, I want to be able to customise the user interface so that it suits the needs of my institution and the repository users As an academic end user, I want sufficient information about research methods to be included with published data, including links to instruments and standards used in data collection and analysis, so that I can reanalyse data and reproduce research results

User stories for funding bodies

g active the second second	-					
Findable	Accessible	Interoperable	Re-usable	Reporting	SLA/Support/Terms	
From a funding body perspective, there would be benefit in an ability to search for a grant applicant's output(s) from previously funded grants, so that compliance with terms and conditions of funding can be verified [ARC]	From a funding body perspective, stable 24/7 access to repositories during the ERA and other reporting periods, would allow that assessors can download research output at all times [ARC]	From a funding body perspective, repository content linked to relevant and authoritative grant information on external websites, would allow end users to better understand the context and conditions of the research [ARC]	From a funding body perspective, repository infrastructure that delivers research findings to business and community stakeholders, would provide greater opportunity for the commercial and social return from research to be optimised [ARC]	From a funding body perspective, the display of metrics relating to research impact and non-academic engagement, would better provide an ability to understand how, where and for what purpose the research is used or applied [ARC]	From a funding body perspective, well-defined support models and SLAs that allow repository content to be secure and accessible for ERA and other assessment purposes would be beneficial [ARC]	
From a funding body perspective, the ability to discover links between funded research and other research projects and outputs, would potentially provide a broader understanding of a research area [ARC]	From a funding body perspective, open access to output of funded research, so that research findings are can be accessed, applied and translated by non-academic stakeholders is important [ARC]	From a funding body perspective, notification of repository content which is not compliant with the policies and conditions of funding, so that exceptions are understood and non-compliance can be recorded would be beneficial [ARC]		As a funder of research I would like to be informed via a report of the progress of the research someone we are funding		
From a funding body perspective, it would be beneficial to be able to easily confirm that repository content complies with funding conditions [ARC]	From a funding body perspective, the display, and active management, of embargo periods for repository items, so that publisher and funding conditions can be monitored and tracked is important [ARC]	From a funding body perspective, notification of access and download statistics for repository items, would allow the impact of OA on funded research to be analysed and understood [ARC]				
From a funding body perspective, repository content that is searchable according to funding schemes and programs, would be useful in identifying research outputs from the	From a funding body perspective, clearly stated rights of use on repository items, would allow end users to confidently and appropriately access and re-use outputs of	From a funding body perspective, the ability to harvest standardised metadata from Australasian repositories, based on funding scheme, FOR codes				

Mapping repository requirements to FAIR

- Findable
 - Extensive and standardised metadata
 - Unique identifiers
 - Comprehensive indexing
- Accessible
 - Non-proprietary internet protocols
 - WCAG compliance
 - Preservation metadata and processes
 - Licences and conditions for access
- Interoperable
 - Common standards and shared languages
 - Machine-actionable metadata
 - Identifiers for everything
- Re-usable
 - Contextual metadata
 - Provenance and rights
 - Research methods

#4: Recommend and make improvements to the current Australasian research repository infrastructure (improve and make the most of what we have).

Analysis of #1-#3 and Recommendations

#1	Australian current infrastructure	Survey and statistical results/analysis
#2	International initiatives	Desktop scan and analysis
#3	User Stories	Phenomenological methodology







Open Access

#1: Of the 45

#2 Repositories are a core part of Open access and Open science infrastructure globally and there are many countries and international organisations/collaborations who have already done substantial investigation/implementation of best practices for repositories. La Referencia, is a Latin American network of open access repositories.

#3: As a publisher I want to ensure that my Open Access policy is adhered to. (Extract from the User Stories Framework)

Analysis:

- Additional user stories (other than the example noted above) related to compliance with institutional policy frameworks or with funder requirements.
- The adoption of institutional Open Access (OA) Policies is significantly less than policies around funder compliance.



Open Access

Recommendation:

In the absence of any national initiatives (e.g. the UK Scholarly Communications Licence or a national Open Access policy) there may be scope for CAUL to provide a checklist of useful elements for an OA policy, either as general statements or as a legal safeguard.



Mapping recommendations to FAIR

- Findable
 - Tags to differentiate between green open access, gold open access, and free to read open access
- Accessible
 - Repositories should be as open as possible and as closed as necessary
- Interoperable
 - Repositories should link data, publications, and grant information as much as possible

• Re-usable

• Provenance metadata to enable appropriate attribution and citation and to help researchers to determine whether they trust the data (for re-use).

Prioritised list of 8 Recommendations

- Recommendation 3: That CAUL set up an ongoing Repository Advisory Working Group. Initial objectives for the CAUL repository technical advisory group should include:
- Recommendation 6: that the proposed CAUL Repository Advisory Working Group advise on recommended minimum requirements for repository systems (including next generation repository systems)
 - *Rationale:* This can be useful when analysing against vendor options with new systems. A checklist would be sufficient.



#5 Ideal State

#FAIR	Selected Statements	When
F.2.	<i>Current</i> All metadata will be made openly available via the relevant repository within three months of publication.	2019
F.3.	Comprehensively open All metadata available via repositories within 3 months. All publications and data within 12 months. ARC will use ERA as an OA driver.	2019
F.6.	Open statistics All Australian repositories will make their usage data, both deposit and usage, openly available. Statistics for OA research outputs will be separated from statistics for digitised library collections and learning objects. Repository statistics will be standardised, such as by using IRUS.	2019
F.7.	Research Australia Collection Australian university research outputs will be findable via a bespoke "Research Australia" comprehensive collection of research outputs including publications, research data and non-traditional research outputs.	2020

#FAIR	Selected Statements	When
A.3.	Funder compliant All Australian universities will be fully compliant with funder policies with respect to availability of metadata, full text open access, and research activity identifiers e.g. ARC and NHMRC.	2019
A.4.	Open Access Policies Universities Australia will have a national approach to Open Access, developed through working with CAUL. All Australian universities will support the provision of Open Access to research outputs via an institutional Open Access Policy, for publications and for data.	2019
A.5.	Green over gold Australian universities are fully Open Access compliant through the provision of green OA via their repository infrastructure, while gold OA via open publication is optional. Metadata and full text of gold OA publications will be deposited into the relevant institutions repository.	2018
A.6.	Affordable Article Processing Charges (APCs) Australian universities will be able to choose to benefit from Publish agreements with publishers, which, separate from Read agreements, and paid separately, will provide the option for authors to pay for APCs at significantly reduced APC costs.	2019

#FAIR	Selected Statements	When
1.2.	Links to grant and funding programs Machine-readable links to grants and funding programs will enable linking between data and publications, supporting reporting to funding bodies.	
1.3.	ORCID Every Australian university researcher will have an ORCID ID to facilitate disambiguation and linking.	
#FAIR	Selected Statements	When
R.3.	<i>Well preserved</i> All Australian university repositories will have a repository preservation strategy and will benefit from active preservation practices.	2019
#FAIR	Selected Statements	When
C.2.	Repository, open access and research management staff are well supported CAUL Repository Community will benefit from a consortium level internationally connected repository technical support group to provide leadership, program management and support.	

#FAIR	Selected Statements	When
C.5.	Be FAIR aware All Australian universities should regularly test the application of the FAIR principles to their repository infrastructure. CAUL should use its annual statistics reports to compile and publish an annual OA compliance report for each institution and the sector as a whole. CAUL should apply the ANDS FAIR data tool to research datasets and publish a report on compliance for each institution and the sector as a whole.	2019
C.6.	Funded infrastructure University and government funding options will be sought to fund the CAUL Next Generation Repositories project.	2019- 2020
C.7.	Contemporary infrastructure Institutions which self-identify as needing a repository software and tools update can benefit from joining the CAUL Next Generation Repository Tools consortium purchase and implementation. Australian repository infrastructure will be built on contemporary best practice repository software tools.	2019- 2020
C.N.	Research Australia Australian [& New Zealand] research outputs will be harvested and made accessible via a Research Australia / Australasia collection and interface built on either NLA Trove infrastructure or on the NGDR platform.	2020-

#6 Repository Tools

Brief summary

Survey of commercially supported tools available.

A number of suitable tools exist.

List of possible requirements.

Enquire of CAUL Council the will to go to market as an opt-in consortium for the procurement of a next generation repository tool.

Work Package Scope

#7 Research Australia

Consider the desirability and feasibility of developing a 'Research Australia' portal/service, providing an entry point to a collection of Australian Research outputs. International initiatives such as OpenAire or SHARE may provide a model.

Phase 1: Establishing Desirability

Semi-structured interviews conducted with 14 stakeholders.

- University librarians, directors, repository managers, grant funding agencies
- Representation from all states, all university classifications (Go8, ATN, RUN)





Phase 1: Establishing Desirability

Key Findings

- 1. Majority of respondents identified some value in having an easily accessible 'picture' of Australian research outputs
- 2. 5 respondents expressed doubt about the value of a new discovery tool/portal (*"Why not just use Google Scholar?"*)
- Majority of respondents agreed any such system should be comprehensive in its collecting (e.g. grey literature, NTRO, outputs from institutions beyond CAUL)
- 4. 4 respondents questioned scope in terms of defining "Australian" research
- 5. No consensus emerged on the mechanisms for delivering such a service

Benefits and Users

Phase 1: Establishing Desirability

- 1. Government, funders, research institutions to showcase national and institutional outputs
- 2. Grant funders as a comprehensive source of funded research
- **3. Researchers** to find collaborators (interestingly no one mentioned to find research)
- 4. Public (including industry) for access to publicly funded research
- 5. Administrators better metrics, benchmarking, process efficiencies

"I support the idea of a Research Australia portal because it gives a national identity."

Barriers and Questions

Phase 1: Establishing Desirability

- 1. We already have Google and TROVE "If we are talking about a shared portal, then I don't see a need for that from either a discovery or preservation perspective"
- 2. the open question only valuable if it provides access to the full text
- **3. Metadata** reliant on the input of quality metadata from existing repositories
- 4. What is 'Australian' policy, technical and marketing questions around what the service would cover

"Should we just focus on Google, or where people are?"

What and How

Phase 2: Establishing Feasibility

- Upgrade TROVE better distinguish/brand Australian research outputs in TROVE and build new features (filters by discipline/licence; grant data; statistics).
- 2. Develop a new portal harvesting repositories the OpenAIRE model
- **3. Shared infrastructure** "There is no sense in 39 of us undertaking the upgrades [to 10 year old repository infrastructure]. We should be looking at a shared collection, shared system for research outputs or data."

Discussion Questions

- 1. Are you interested in a Research Australia service?
- 2. Who do you think would use the service?
- 3. Have you encountered any use cases that would have benefited from Research Australia?
- 4. What would be the preferred model:
 - a. Enhanced TROVE
 - b. New portal (e.g. OpenAIRE)
 - c. Shared repository infrastructure
 - d. Other

www.menti.com (if we have time)

Contact: belinda.tiffen@uts.edu.au

OR speak to a member of the team today!

Also bring to your attention:

- Project Initiation Document
- Project Plan
- Comments from CAUL Council 2018/2
- Finalised papers will go to CAUL Council 2019/1 for resolution and approval.
- A new project team will be formed to progress implementation.

Draft Recommendations

- 1. Endorse the Review and report on the current Australasian institutional research repository infrastructure.
- 2. Endorse the review of international research repository infrastructure and developments.
- 3. Seek specific project funding to develop required national/regional repository infrastructure.
- 4. Seek a consortial membership of COAR.



- 5. Establish ongoing repository technical advisory working group
- 6. Establish a group which reviews the training and professional development required for repository staff.
- 7. Endorse the report on repository user stories.
- 8. Endorse the user stories which inform development of future Australasian repository infrastructure.



- 9. CAUL endorse the report from Product 4: Recommend and make improvements to the current Australasian research repository infrastructure (improve and make the most of what we have).
- 10. That the proposed CAUL Repository Advisory Working Group advise on minimum metadata standards.
- 11. That the proposed CAUL Repository Advisory Working Group develop an Open Access Policy template.
- 12. That the proposed CAUL Repository Advisory Working Group advise on recommended minimum requirements for repository systems.



- 13. That the proposed CAUL Repository Advisory Working Group provides advise related to metadata standards and technology for sharing across Open Access platforms and standards.
- 14. That the proposed CAUL Repository Advisory Working Group liaise with ARDC about institutional requirements for training.
- 15. Endorse the report on the ideal state for Australian repository infrastructure and FAIR access to research more generally.
- 16. Endorse the statements around the ideal state for Australian repository infrastructure and FAIR access to research.



- 17. Endorse the statements on ideal state and these be socialised amongst other stakeholder groups including the Australian Research Data Commons (ARDC), Australian Research Council (ARC) and National Health and Medical Research Council (NHMRC).
- 18. Endorse the list of repository tools.
- 19. Endorse the general repository requirements.
- 20. Consider the pros and cons or a consortial approach to repository procurement.



- 21. Consider progressing a consortial approach to the procurement of a next generation repository system.
- 22. Develop a project cost sharing basis and fund the project, if applicable.

Additional recommendation:

21. Support an extension for project #7 Research Australia Collection to December 2018 to facilitate the completion of this work.





Group work at your tables

 Provide responses to the report and the recommendations using butchers paper -















