CAUL
COUNCIL OF AUSTRALIAN UNIVERSITY LIBRARIANS

CAUL Fair, affordable and open access to knowledge program

CAUL Review of Australian Repository Infrastructure

Review of Australian Repository Infrastructure

A project within the CAUL Fair, affordable and open access to knowledge program

To: Director, FAIR Access to Research Program, CAUL

From: Project Leader, Review of Australian Repository Infrastructure

Subject: Report from the Review of Australian Repository Infrastructure

Date: Version 15 March 2019

Dear Catherine,

On behalf of the Project Team I am pleased to provide you with the report from the CAUL Review of Australian Repository Infrastructure, a project of the CAUL Fair, affordable and open access to knowledge program.

The project was resourced using a large number of volunteers from across the CAUL member libraries. This model has proved to be successful.

This final version of the report includes reports from work packages #1 - #7. All feedback provided by the 2/2018 CAUL Council meeting a community consultation process conducted in February/March 2019 were considered and incorporated into the document where possible. Suggested changes to project scope could not be incorporated into the project and the report at this stage.

Best regards,

Martin Borchert
Project Lead

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Table of Contents

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Project Overview

Project Initiation Document

The original project initiation document, produced by Jill Benn, is provided as Appendix 1.

Project Plan

The project plan, produced by Martin Borchert and refined by the project Steering Committee, is provided as Appendix 2.

Project Steering Committee Members

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

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Executive Summary

The purpose of the CAUL project is to review the current state of the Australian institutional research repository infrastructure. The objectives were to make observations and recommendations on possible ways forward for the Australian repository infrastructure and management.

CAUL executives approved a Project Initiation Document (PID) for CAUL Review of Australian Repository Infrastructure Project in late 2017 and called for EOIs to lead the project. Martin Borchert, University Librarian at UNSW Sydney, was appointed to lead the project.

An EOI was sent out for community members to join the project. A Project Steering Committee was established in early 2018 and a Project Plan was developed which set out 7 work packages to be completed to achieve the project goals. A work schedule was established which ensured work packages would be carried in parallel and making reference to each in turn, in order to achieve producing outputs in time for the CAUL 2018 / 2 meeting in September.

Information was gained through discussions amongst project team members, consultation with stakeholders via an online survey and semi-structured interviews, and by assessing repository infrastructure, processes and policies against the FAIR principles – Findable, Accessible, Interoperable and Reusable.

A draft project report consisting of work packages #1 - #6 was presented at the CAUL 2018 / 2 meeting and feedback was sought.

Work package #7 was completed in January 2019 and added to the draft report.

A community engagement process was conducted in February – March 2019 and consisted of a Zoom online video session for community members and a SurveyMonkey online form used to solicit feedback against each work package and recommendation. Feedback is provided in the Appendices.

Feedback was considered by the Project Steering Committee and the final report produced for presentation to the CAUL 2019 / 1 meeting (April).

Through the completion of work packages, it was found that:

- Australian repository infrastructure is diverse with numerous mature and new generation, local and cloud hosted, open source and proprietary solutions in use.
- While all institutions endeavour to make published research outputs openly available, only half make research data available.
- Non-research related collections – archival library collections, images and multimedia, and course materials are also in scope for a minority of institutions.
- Less than half of Australia's universities have an Open Access Policy or statement for research outputs.
- Two-thirds of institutional repositories support grant funder policies, only one-third monitor compliance, and one-quarter use Research Activity Identifiers.
- Only one-third of institutions have a preservation strategy for their repository collection.
- While [lack of] harvesting by Google and Google Scholar was initially thought to be an issue, this is not the case with 90% of institutions reporting satisfactory harvesting. Further inspection demonstrated that the remainder of institutional repository

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

collections are actually being harvested either directly, or via Trove which is harvested by Google, but not Google Scholar.

- Interoperability is of primary concern to the current and next generation repository systems.
- The Australasian repository landscape would benefit from greater connection with international repository developments such as via Coalition of Open Access repositories (COAR) membership.
- The Australasian repository landscape would benefit from increased coordination and support via a technical advisory group which could lead initiatives on training, interoperability, metadata, standards and system requirements. This group would build upon the excellent work undertaken by the CAIRSS group / CAUL Repositories Community Group.
- User stories based around Findable, Accessible, Interoperable and Reusable can be defined, which help to direct resources and attention for repository development.
- Existing Australian repository infrastructure can be improved via the work of a technical advisory group, however the diversity of systems and investment in the landscape, means that this is complex and that benefits would not be evenly spread across institutions. Many institutions report a reluctance to further invest in their legacy repository system, having a preference for a next generation system.
- A series of ideal state statements could be identified to provide direction to the technical advisory group and future project work. Again, the diversity of institutional policies and infrastructure makes it difficult to apply ideal state statements across the landscape.
- A range of attractive, current and next generation repository tools are available, these being open source, proprietary or commercially hosted and supported open source. This suggests an opt-in consortium procurement and implementation process may be a suitable avenue for updating institutional repository systems to bring them uniformly up to minimum system requirements and standards. It was found only a minority of institutions, perhaps a hand full, would partake in a shared infrastructure procurement process, questioning the financial viability and sustainability of such a project. This was confirmed by the consultation process. Shared infrastructure would likely be on a cloud provider platform.
- The success of international repository networks and collections may point the way for the development of same in Australasia. While a “Research Australia” collection was initially thought to be of interest to the sector, this was not found to be the case, with little interest in and support for developing such a collection, especially considering likely project costs to the sector. This was again confirmed by the consultation process. The NLA Trove system is the closest thing Australia has to such a collection and infrastructure. Working with the NLA to separate the national research collection from the library collections and developing a Research tile in Trove, and ensuring indexing via Google Scholar, offers the least work-intensive and most cost-effective solution to achieving this, while also being independent of any shared repository infrastructure project.
- The Project Initiation Document (PID) scope intentionally excluded the investigation of Current Research Information Systems (CRIS) / Research Management System(s) and their interoperability and integration with research repositories. Feedback gained from the consultation processes indicates some community members regards this to weaken the value of this report.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Project Products and Teams

The project was designed with the following seven project products. Teams were required to collaborate and undertake work in parallel as much as possible in order to achieve projects goals within the prescribed timeline.

Project products

1. Review and report on the current Australasian research repository infrastructure.
2. Review and report on the international research repository infrastructure and developments.
3. Develop a series of repository user stories.
4. 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.
6. Investigate and make recommendations for next generation repository tools (consortium approach, possible infrastructure project).
7. Investigate and make recommendations for a possible “Research Australia” collection of research outputs.

Project working space

Project documents folder on Google Docs

https://docs.google.com/drive/folders/1wMVkXg0YG_RoMeFCKU_mznXoibVog4qo

Please contact Martin Borchert or Harry Rolf to join the Google Docs group.

Consultation

Community consultation and feedback was incorporated into the project throughout and led to many refinements of the report.

- The project team included representatives from across CAUL member libraries.
- The draft report and presentation slide summary were provided to CAUL Council 2/2018 where feedback was provided by University Librarians. (Appendix 8)
- The draft report and presentation slide summary were provided to the CAIRSS / CAUL Repository Community Day attendees in 2018.
- A Consultation webinar and online survey were provided to CAUL members and CAIRSS group in February/March 2019 (Appendix 9).

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

LIST OF KEY RECOMMENDATIONS

The Product #1-7 reports presented the following key recommendations that CAUL:

1. Seek to increase representative consortial membership of COAR by CAUL/CONZUL libraries.
2. Establish ongoing repository technical advisory working group which will oversee training, minimum metadata standards and repository system requirements and open access policy template and liaison with partners including National Library of Australia (NLA), Australian Research Data Commons (ARDC), Australian Research Council (ARC) and National Health and Medical Research Council (NHMRC).
3. Consider the pros and cons for a consortial approach for shared repository infrastructure, given limited interest from the sector.
4. Develop a project cost sharing basis and fund the project, if applicable.
5. Work with the NLA to develop a Research Outputs tile in Trove as the most cost-effective pathway to providing a comprehensive Research Australia style collection and to ensure all outputs are harvested by Google Scholar.

All recommendations are provided at the end of each work package #1- #7.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

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

Authors

Andrew Harrison, Research Infrastructure Librarian, Monash University (Lead)

Maude Frances, Associate Director, Library Repository Services, UNSW

Mary O'Connor, eResearch Support Project Librarian, University of Adelaide

Kay Steel, Manager, Research and Strategic Projects, Federation University Australia

Introduction

This report is part of the Fair, Affordable and Open Access to Knowledge Project led by the Council of Australian University Librarians (CAUL). The objective of this project is to determine how improvements to repository infrastructure can be made across the sector to increase findability, accessibility, interoperability and reusability (FAIR) of Australian-funded research outputs across the sector.

The project approached this large task by creating several working groups each focused on specific objectives of the wider project. This Review and report on the current institutional research repository infrastructure in the Australasian universities sector is the report for work package #1.

Methodology

The principal tool for gathering the data analysed in this report was a survey of 48 institutions in the Australasian region. The survey was conducted through May - June 2018 with 45 universities responding, including all eight of the New Zealand institutions, and all but three of the Australian institutions, which is a total response rate of 94%.

There is evidence of under reporting in the survey, based on the authors' joint knowledge of the number of some repository software instances actually in use in Australasia, when compared to the number reported by the survey respondents; so, whatever the cause of the incomplete data provided, some caution should be applied to the data. Nevertheless, and unless otherwise specifically noted, the report was analysed on the data presented and not on the data inferred.

The review did not repeat the work done on the recent CAUL 2017 survey[1]. The content of that survey report is recent enough to still be relevant and should be read in conjunction with this report, especially for information on staffing levels, identifiers, metadata, article processing charges, promotion/feedback, and discovery.

Scope of Project

- Only CAUL and CONZUL institutions were surveyed.
- Answers were analysed against the FAIR principles.
- No significant overlap with the data collected in the 2017 CAUL survey.
- Additional information was gathered to inform other aspects of the wider project.

Findings

This section presents the detailed findings and some analysis from the questionnaire, occasional comparison with the recent CAUL 2017 survey, and work member comments based on team knowledge. The relevant questions asked in the survey are listed under

[1] Simon Huggard, Kay Steel, Alexander Sussman. Research Publications Repository Survey Report 2017 - Dec 2017. http://www.caul.edu.au/news/research-publications-repository-survey-report-2017

each of the section headings. Analysis of the FAIR principles[2] is discussed in detail under the section "Self-assessment of the FAIR principles" (pp. 11-12).

Repository infrastructure and content

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

Infrastructure

Figure One lists the repository software by percentage of institutions using it. The graph was limited to the more popular version of repository, leaving off the single use instances.

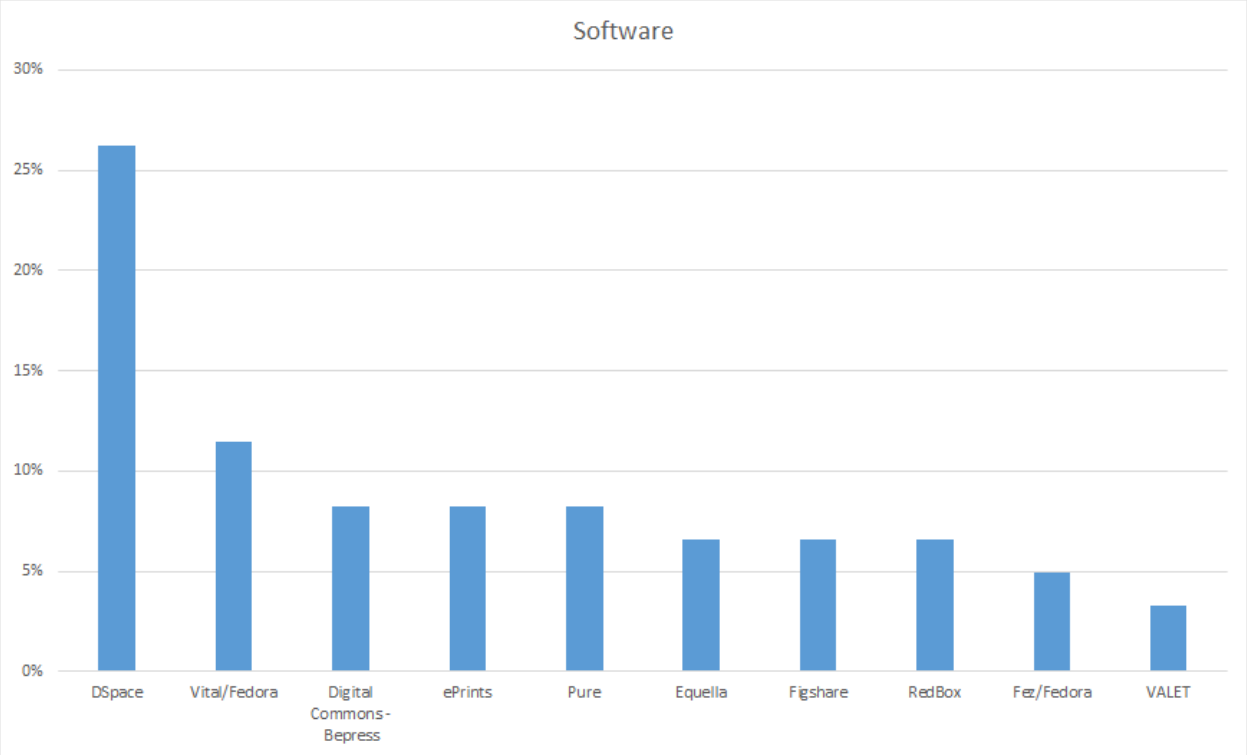


Figure 1: Percentage of institutions using repository software.

Australasian universities used DSpace (16 universities) and Fedora based repositories (12). The next level of commonly used software is ePrints (5), Digital Commons/Bepress (5) and Pure (5). Other software includes RedBox (4), Equella (4) and figshare (4). One respondent uses Alma/Primo/Appian and another reported using an Omeka repository.

While the question of the age of the infrastructure was not specifically asked, the authors of the report could identify that nine institutions had newer generation repository software in their infrastructure descriptions. The number of different repositories instances reported by each institution were that 25 universities used one repository, 14 used two and four used more than two repositories. When counting software types specifically, 40 universities use

[2] The FAIR Data Principles - accessed 31 July 2018. https://www.force11.org/group/fairgroup/fairprinciples

more mature software such as DSpace, and of these a small number reported they are transitioning to new software, such as Pure and figshare. This implies that most repository infrastructure is mature and this is perhaps due to the fact that Federal Government funding for institutional repositories, the Australian Scheme for Higher Education Repositories (ASHER) which began in 200,7 is now 11 years in the past, with no further funding to allow institutions of all sizes and capabilities to keep their repository infrastructure update with modern capabilities and new standards such as ORCIDs. The authors are aware of software that is in use but not mentioned in the survey. For example, there are seven instances of figshare in operation but the survey reports only four instances.

The eight New Zealand (NZ) universities use DSpace, while the five Australian Technology Network (ATN) respondents utilise a range of software. The Innovative Research Universities (IRU) use a range of systems, while the Regional Universities Network (RUN) mostly use Fedora. In the Group of Eight (Go8), five universities use two or more systems, with four instances of DSpace and two inhouse offerings. The private universities both use Bepress. The remaining nine unaligned universities use Bepress, Equella, Eprints and Fedora.

Overall, the software breakdown is similar to that recorded in the CAUL Research Publications Repository survey 2017 for nearly all types. Market share for DSpace and VITAL has dropped by three institutions apiece. Figshare has entered the market with several institutions using or in the process of adopting this product.

Content

All universities use their repositories to house research outputs/publications, and thirty-one state that they store theses there. The actual number of repositories storing theses is probably higher, and some respondents may have assumed that the terms “research outputs” or “research publications” included theses.

Datasets were reported in the collections of twenty-two universities, which is less than half of the total respondents and perhaps reflects the broad nature of the infrastructure question leading to under-reporting of what is actual practice. Archival and other library collection material was part of the collections of sixteen universities. Five were hosting university journals, and three displayed images/multimedia files.

Eight universities listed other content, mostly course readings and other learning and teaching artefacts.

Seven of eight NZ universities report collecting theses in their repository infrastructure. Two collect data, one hosts journals and one hosts the New Zealand electronic text collection. Three of five ATN universities mention theses, while only one collects data via the repository. Five of the seven IRU universities mention theses, and three universities display archival and/or library collections. Four collect datasets, and one mentions learning and teaching resources. Of the Go8 universities, seven mention theses. Seven have archival and/or library collections, and six collect data. Three mention other content, mostly learning and teaching material. Of the five RUN respondents (of six RUN universities), three mention theses, while all five collect data. One hosts images, while another hosts journals and conferences.

The two private universities and the remaining nine unaligned universities mostly mention theses, and five have archival and/or library collections in their repositories. Three collect data, two host journals and two mention other content.

Open Access

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

Of the 45 respondents, 16 have an institutional Open Access (OA) Policy; 15 have a partial OA Policy, and 14 have no OA Policy. 20 respondents expect change to the current situation within the next 12-24 months, with this spread across institutions which already have an OA Policy (six expect change), a partial OA Policy (nine expect change), or without a policy (five expect change). There is no apparent correlation between the age of the repository software in use and an expectation of change in institutional OA Policy.

All eight New Zealand universities submitted a response. Of these, three have an institutional OA Policy, one partial and four do not have an institutional OA Policy. Two respondents currently without an institutional OA Policy expect this to change in the next 12-24 months.

All Go8 universities responded. Of these only one had no OA policy, although this institution indicated that change was likely. Two institutions had OA policies and five had partial policies. Change was expected in one of the former and four of the latter respondents.

Among the six RUN universities, of five respondents three had no OA Policy (although one expected change), and two had partial OA Policies.

Of the two private universities, one has a partial OA Policy and expects change.

All of the seven IRU network responded, with three having neither a full nor a partial OA Policy, but five universities expecting change.

Of the other 14 respondents, only one had neither a full (eight) or partial (two) OA Policy; six expect a change in the next 12-24 months.

The overall response is similar to that received in the CAUL Research Publications Repository survey 2017, when 36% of universities mandated full deposit of research publications into repositories and this survey found 37%.

An examination of the policies linked to/referred to by the responding institutions highlighted a variation between definitions of ‘full’ or ‘partial’ policies. There was some variance here, with a number of the ‘partial’ policies providing greater expectations around either publication of research outputs via an OA mechanism, and/or deposit of outputs to an OA repository, than some of the declared ‘full’ OA policies. From this it can be suggested that the landscape around OA policies is more comprehensive than a straight numerical tally of responses might suggest.

An analysis of those responses of ‘full’ OA policies indicates that the consensus is for the required deposit of the ‘Accepted’ version, possibly in addition to the published version, to the institutional repository. In the overwhelming majority of policies, thesis submission is also

mandated. One policy specifically excludes research data; while another specifically includes non-traditional research outputs.

An analysis of the responses relating to 'partial' OA policies indicates wording such as 'encourage' OA publication, or OA is supported 'in principle', or 'encourage OA deposit' of research outputs, or 'where possible' appropriate versions of research publications should be deposited. Again, the overwhelming suggestion is for green OA via deposit to the institutional repository, although some policies also mention discipline-specific repositories as an acceptable alternative. Theses are often mandated for deposit. Research data is included in five policies, and HDR research data is specifically included in one of these.

In summary, institutional policies around OA did not usually specifically refer to research data; HDR thesis deposit was generally required (although limited embargo periods were sometimes available); and the primary means of OA access referred to was via the institutional repository although a discipline-specific alternative was also suggested in some policies.

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?

Grant funder policies were supported by 33 institutions, with two institutions answering no and a further ten institutions giving a qualified answer of other. The qualifications were mostly that the repository supported the Grant funder policies passively, by virtue of the requirement to support depositing of open access research outputs in general or NHMRC and ARC grants in particular; confusion as to what the question was asking, and, in one case, because the process of supporting grant funder policies was completely manual. Converting a number of these 'other' answers into a 'yes', because any support, manual or selective, counts, the total 'yes' responses rose to 95%. One respondent noted that a lack of ORCID integration in their repository may impact their compliance with the latest NHMRC OA policy[3], which mentions ORCIDs specifically as a feature of a compliant repository record.

Monitoring for compliance with the policy was reported positively by 14 institutions. Five of the no answers were actually qualified yes answers because the monitoring was not yet systematic, complete or was managed by another part of the university. This now totals to 42% of universities doing something to monitor their own compliance with grant funder OA policies. The compliance process was managed by the library in nine responses and by the research office in five. Three respondents also reported that their institution was thinking about monitoring compliance. There was no apparent pattern to why they did or did not monitor compliance with grant funder policies.

[3] See Section 10 of Appendix, NHMRC Open Access Policy 15 January 2018, https://www.nhmrc.gov.au/_files_nhmrc/file/research/nhmrc_open_access_policy_15_january_2018_v2.pdf

Accessed 11/07/2018.

Research Activity Identifiers

Question: Does your repository use Research Activity Identifiers.

Only five of the responding institutions stated that they used or stored research activity identifiers, like RAiDs, in their infrastructure and systems. None of the yes answers actually mentioned using the RAiD identifier itself, instead a variety of other identifiers were reported. These were mostly ORCIDs, handles and DOIs, that all identify things related to the research outputs, people and publications, but not the research activity itself. Only one respondent reported that their infrastructure supported using RAiDs but its use had not yet been implemented. The majority of the 'Yes' respondents (four) are part of the Group of Eight network and they used DSpace, Fedora and Pure repositories.

Repository Interoperability and Integration

Question:

Is your content harvested by any of these services?

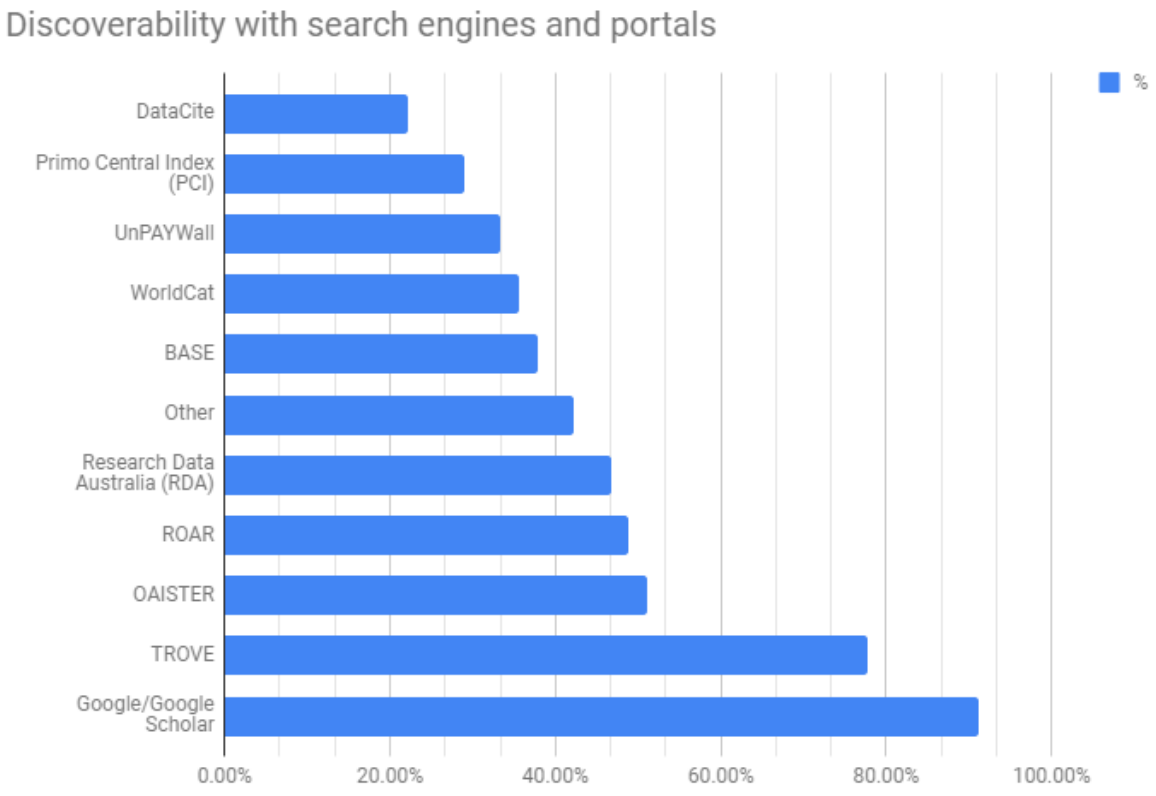


Figure 2. The proportion of institutions reporting harvesting by certain search engines and portals.

Questions:

a). Describe the integration of your infrastructure with other systems both internal and external.
b). How is metadata technically shared with other systems e.g. OAI-PMH, SiteMaps and custom APIs.
c). Do you actively monitor harvesting pathways? e.g. Do you regularly test harvesting outcomes with key targets.

When asked to select from the ten named search services or portals listed in Figure 2, 25 institutions selected five or more dissemination targets, nine institutions selected seven or more targets, and three institutions selected eight, nine and ten targets respectively.

Google, Google Scholar and TROVE were identified as the most common harvesting targets by over 70% of the respondents. OAISTER, ROAR and RDA were the next most popular group with between 45% to 50% of respondents naming them. After the catch all category of 'Other', the next level of popularity was BASE, Worldcat and Unpaywall, with 30% to 40%. The high number of Australian contributors to TROVE, nearly on par with the obligatory Google indexing, is evidence of strong interest in having a national portal for the institutional repositories' combined research outputs.

The respondents were then invited to document other harvesting agencies and services that they served in the 'Other' category. This data shows 18 institutions had one or more additional harvesters listed, but only three of those institutions had three or more harvesters, and only one institution named up to four other harvesters.

In the 'Other' category four respondents named CORE, three each named Open Access Thesis Dissertations (OATD), NZResearch, two named Digital NZ and Summon, and there were single mentions of Research Data Australia (RDA), Research Gate, RePEC (Research Papers in Economics), Primo Central, Digital Commons and Open Dissertations.

All respondents answered that they were using the Open Access Initiative - Protocol for Metadata harvesting (OAI-PMH) to share metadata records. Sitemaps and Application Program Interfaces (APIs) were the next most common technology used to integrate systems and share data.

The most common systems integration reported was with the institutions' Research Information Management (RIM) systems, often in conjunction with the profile management systems. The most popular external system was the ORCID system. Only one institution disclosed that they had no integration with systems outside their repository system.

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,

a). SSRN: Social Science Research Network

b). PubMed central

c). APO

d). Others

20 respondents answered no to this question or 44% of the answers. Scopus and Web of Science were defined as discipline repositories by a number of the respondents, with

REPeC, SSRN, PubMed central APO and arXiv.org being commonly named by the other respondents.

Strengths and weakness of repository infrastructure/environment

Questions:

What do you consider are the strengths of your repository infrastructure/environment?
What are its weaknesses - what could be changed/enhanced/done better?

When reviewing the self-assessment of the strengths and weaknesses of the respondent's repository infrastructure, the following sets of words were noted by the authors as frequently mentioned in the answers.

Integration was most often mentioned as a strength and a weakness in the same set of answers from an institution, reflecting the particular nature of integration challenges that a repository infrastructure had with internal and external data providers and clients in other systems. Interoperability was picked out as a common strength by several of the respondents, with one answer highlighting that the repository was integral to the research management of the respondent's university.

Robust or stable was another high frequency set in the answers given for the strengths, indicating a desire for repository software that was not in need of constant attention and intervention to maintain operational functionality.

Less frequently mentioned but still significant enough to be noticed in a read through of the answers were the following:

open source vs proprietary systems,

technology lifecycle management, vendor and system lock-in and cost of change,

cloud and hosting,

mature software in the context of a weakness, lack of development future,

preservation capability,

automation as both strength and the lack of as a weakness, and

identifiers, mostly in the context of author identifiers like ORCIDs, was also both a strength and a weakness.

Preservation strategy

Question:

Do you have a preservation strategy for the repository content?

Is it an institutional or a library policy and can you provide a link to the document?

Otherwise is there an informal commitment or understanding?

Less than one third (13) of the respondents answered that they had a digital preservation strategy for their research repository content and three institutions reported that they are developing a strategy. Six of these institutions either declared, or linked to, a preservation policy document.

Ten institutions across both the Yes and No answer groups reported having separate backups of repository data. Based on the descriptive answers supplied, the backups appeared to be the only preservation activity for some organisations. Format preservation was mentioned by three institutions, both in choosing an archival format, such as PDF(A), and format validation and preservation metadata, such as JHOVE and PREMIS. Another three institutions reported that the vendor managed the preservation of the repository content on their behalf.

Self-Assessment of the FAIR principles

Question: Consider your repository infrastructure in relation to the FAIR principles.

In general, repositories supported the FAIR principles for findability and accessibility more than those for interoperability and reusability. A number of institutions plan to consider FAIR principles more closely in future repository implementations and several stated that they did not have resources to align their repository infrastructure with FAIR principles at present. It is important to note that comments about alignment with FAIR principles depended on the respondents' understanding of the principles. It was clear from the responses that levels of understanding varied greatly across institutions.

Inclusion of PIDs was the most commonly stated indication of findability. Handles were most frequently mentioned followed by DOIs, ORCIDs and identifiers of funding bodies. Several institutions proposed that while inclusion of DOIs and ORCIDs were priorities, implementation was not currently possible due to either software or vendor constraints. Rich and standardised metadata was proposed by several institutions as evidence of findability. Frequency of comments relating to harvesting of repository content by browsers and inclusion of content in national and international databases were consistent with responses to an earlier question about services which harvest repository content, according to which almost two thirds of the repositories provided content to more than five external services. Several institutions reported that their records were well indexed by Google and Google Scholar.

There was high variability on self-rated alignment with the FAIR principle for accessibility. A major focus in responses was on the proportion of repository content which is openly accessible, with little reference, beyond use of standard file formats and Handles, to the technologies that enable the access. While a small number of repositories reported that, except for embargoed content, their repositories were completely OA, many reported that less than a third of their records were openly accessible. This was thought to be due more to policies and user behaviour than to technical limitations of the infrastructure. A few repositories mentioned request-a-copy buttons or manual procedures for mediated access to content, and one repository has implemented NISO tags to support machine-to-machine recognition of OA content. Several respondents suggested that licence information in their

repositories was not always clear and that this was an impediment to better alignment with the accessibility principle.

Of the third of respondents who considered that their repositories aligned with the interoperability principle, more than half suggested that the alignment was only partial. While Dublin Core (DC) and other standardised metadata enable some degree of interoperability, there was awareness that protocols for interoperability have shifted since many repositories were implemented over the last decade or so, and that existing software did not enable adoption of FAIR requirements for controlled vocabularies and machine-readable formats. In some cases, repository-specific interoperability was constrained by dependence on the institution's publications management system as a data source. DOIs, Handles and ORCID identifiers were mentioned by some respondents, as well as general adoption of community agreed formats and standards and vocabularies including FOR codes and grant identifiers. While a few respondents referred to links or qualified references to other meta(data), others noted the limitations of DC for articulating relationships between entities.

The FAIR principles relating to reusability were not well addressed in repositories at most institutions, and several respondents reported that because their repositories did not hold research data the principles were not relevant for them. Several respondents reported that while provenance and rights information is included in descriptive metadata of their repositories, licences are generally not machine readable and contextual information such as version details may not be accessible to end users.

Community feedback indicated the project could have defined the distinguishing features of next generation repository systems and then later applied these criteria to an assessment of the numerous systems used and available. This was not included in the project plan and was not carried out post-project. Instead a reference is provided to the Coalition of Open Access repositories (COAR) Next Generation Repositories project site https://www.coar-repositories.org/activities/advocacy-leadership/working-group-next-generation-repositories/

Recommendation

1. That the CAUL Project: Review of Australian repository infrastructure group accepts the Review and report on the current Australasian institutional research repository infrastructure.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

#2: Report on International research repository infrastructure and developments

Authors:

Ginny Barbour, Director, AOASG (Lead)

Natasha Simons, Program Lead, Skills Policy and Resources, ANDS

Scott Abbott, Manager, eResearch, eScholarship, UTS

Alexander Sussman, Associate Director, Academic Services, UNSW

Executive Summary

Institutional repositories are core to Australia's library and research infrastructure and are also key to supporting open access. The need for a review of the Australian repositories landscape has been identified by CAUL as part of its Fair, affordable and open access to knowledge program. In order to provide an international context we undertook a review of relevant international research repository infrastructure and developments.
In this review we provide a summary in Table 1 of the key infrastructure, tools and supporting organisations that we assessed (further details are provided in Appendix 1.

We identified the following key themes from the review:

- Overall there was increasing complexity across the sector.
- The most important characteristic of any initiative and where most effort is being expended is for global interoperability.
- There is a high degree of interdependence between successful initiatives and at its best there is close collaboration.
- Securing stable funding is essential for any successful initiative.
- Although the global context is important, regional and national networks have a key role.

Finally, we note that 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.

Based on this review we make five recommendations:

1. CAUL endorse the review of international research repository infrastructure and developments.
2. CAUL seek specific project funding to develop required repository infrastructure.
3. CAUL seek a consortial membership of COAR.
4. CAUL set up an ongoing, professionally supported, repository technical advisory working group.
5. CAUL set up a group which reviews and coordinates training and professional development required for repository staff.

Background

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. As Australia takes the opportunity to review its own required infrastructure, it is thus timely to review what is happening internationally. This review is not intended to be a wholesale review of the history of repositories globally, nor is it an in-depth review of specific national or subject based repositories, except when they serve to illustrate a specific point. What this review specifically sought to discover, and review were examples of robust ongoing structure or emerging best practice in any aspect of repository development. In the context of a complex

international situation,[4] the aim was to provide timely, relevant recommendations for the Australian context.

We note that in the past couple of months there have been two important international repository meetings - the COAR[5] and OR2018[6] - meetings and discussions at these meetings are of high relevance here.

Methodology

In compiling this report, we began by brainstorming on known examples of good practice among the group. Our initial source of examples was derived from work in 2017 done by the previous AOASG and CAUL repository working group led by Ginny Barbour, AOASG and Natasha Simons, ARDC.

The examples to be included were refined through an iterative process through discussion within the group. Detailed information on relevant initiatives and specific practices was compiled by members of the group from the relevant websites, including from more general sources of information such as COAR[7], ROARMAP[8], and where needed from personal contacts at the initiatives.
We classified the examples into three broad categories of "Infrastructure", "Tools" and "Supporting Organisations". Where appropriate, examples are noted as international / transnational (the predominant category) or national.

We took a structured approach in assessing each of the examples that we identified and extracted the information for each into a table. In this summary document we list the examples reviewed (Table 1). In the sections following we pull out the key messages. The full information that we extracted for each example are in Appendix 3.

Table 1: Infrastructure/Tools/Supporting Organisations reviewed

Further details for each initiative are in the Appendix 3.

Name (Location)	Infrastructure/ Supporting Organisations/ Tool	Short description/notes

[4] Arlitsch, Kenning, and Carl Grant. "Why So Many Repositories? Examining the Limitations and Possibilities of the Institutional Repositories Landscape." *Journal of Library Administration* 58, no. 3 (March 2018): 264–81. https://doi.org/10.1080/01930826.2018.1436778.

[5] "COAR » COAR Annual Meeting 2018." Accessed June 18, 2018. https://www.coar-repositories.org/community/coar-annual-meeting-2018/.

[6] "Open Repositories 2018." Open Repositories 2018. Accessed June 18, 2018. http://www.or2018.net/.

[7] https://www.coar-repositories.org

[8] https://roarmap.eprints.org/

La Referencia (Latin America) http://lareferencia.redclara.net/rfr/	Infrastructure	La Referencia, is a Latin American network of open access repositories. Through its services, it supports national Open Access strategies in Latin America through a platform with interoperability standards, sharing and giving visibility to the scientific production generated in institutions of higher education and scientific research.
OpenAIRE (Open Access Infrastructure for Research in Europe) https://www.openaire.eu/	Infrastructure	OpenAIRE is a massive and successful infrastructure project that is well funded by an organised EU. Lessons can be learnt from the centralised top down funding and management of such a complex and successful project. RIOXX to be folded into the Open AIRE metadata schema.
Open DOAR (International) http://www.opendoar.org/	Infrastructure	*Open*DOAR is an authoritative directory of academic open access repositories.
PubMedCentral - PMC (US) http://www.opendoar.org/	Infrastructure	PMC is an American free archive/repository for biomedical and life sciences journal literature deposited by participating journals, as well as for author manuscripts that have been submitted in compliance with the public access policies of participating research funding agencies.
SHARE - SHared Access Research Ecosystem (US) http://www.share-research.org/	Infrastructure	As an established platform with back-end and front-end source code (for the research database and the platform) open source, SHARE is a structure worth looking at more closely. SHARE adapts to the data sources' metadata structure and to refine this process they have recently published the SHARE Version 3 Metadata Harvesting update. Current apparent lack of connection between Open AIRE and SHARE, however, needs to be considered.
CORE (UK) https://core.ac.uk/	Tool	CORE (Connecting Repositories) is a UK based free service aimed at aggregating all open access content distributed across different systems.
Dataverse (International) https://dataverse.org/	Tool	Dataverse is an open source web application to share, preserve, cite, explore, and analyse research data. A Dataverse repository is a software installation, which then hosts multiple virtual archives called Dataverses. Each dataverse contains datasets, and each dataset contains descriptive metadata and data files (including documentation and code that accompany the data).
Dryad (International) https://www.datadryad.org/	Tool	Dryad is an international disciplinary repository for data underlying scientific and medical publications. Dryad is open source DSpace repository software

		and advocates for making data Findable, Accessible, Interoperable, and Reusable.
Figshare (International) https://www.datadryad.org/	Tool	Figshare is a privately owned company. Cutting edge proprietary system to surface open research content (data and publications) from a wide variety of sources. Very attractive in terms of a readymade solution. Their publisher model requires an SLA statement guaranteeing (only) 10 years of persistent availability. (From here)
IRUS-UK - Institutional Repository Usage Statistics United Kingdom (UK) http://www.irus.mimas.ac.uk/	Tool	A repository plugin which provides COUNTER-conformant usage statistics for all content downloaded from participating UK institutional repositories (IRs). There is currently an Australian trial underway.
Re3data (Registry of Research Data Repositories) http://re3data.org/	Tool	re3data.org is a global registry of research data repositories from different academic disciplines. It is worth noting that re3data will be used in the AGU FAIR project as the basis of a tool researchers can use that guides them to select which repository to best deposit their data in.
ResourceSync https://www.niso.org/standards-committees/resourcesync	Tool	ResourceSync is an ANSI/NISO specification that describes a synchronization framework for the web consisting of various capabilities that allow third-party systems to remain synchronized with a server's evolving resources.
RIOXX -Metadata Application Profile United Kingdom (UK) http://rioxx.net/	Tool	RIOXX is a metadata application profile that provide a mechanism to help UK institutional repositories comply with the RCUK policy on open access.
Scholix initiative (Scholarly Link Exchange) (International) http://www.scholix.org/	Tool	Scholix is a work in progress from Research Data Alliance; model to link publications with data; will have global application.
CARL - Canadian Association of Research Libraries (Canada) http://www.carl-abrc.ca/advancing-research/institutional-repositories/repos-in-canada/	Supporting Organisation	Peak body for Canadian academic libraries.
COAR -Confederation of Open Access	Supporting Organisation	COAR is an international association with over 100 members and partners from around the world

Repositories (International) https://www.coar-repositories.org/		representing libraries, universities, research institutions, government funders and others. COAR brings together the repository community and major repository networks in order build capacity, align policies and practices, and act as a global voice for the repository community.
LERU (Europe) https://www.leru.org/	Supporting Organisation	The League of European Research Universities (LERU) is a well-established network of research-intensive universities.
LIBER (Europe) http://libereurope.eu/	Supporting Organisation	LIBER (Ligue des Bibliothèques Européennes de Recherche) - is a European network of more than 400 national, university and other libraries.
National Institute of Informatics (NII), (Japan) https://www.nii.ac.jp/en/	Supporting Organisation	The NII is an inter-university Japanese research institute for advancing the study of informatics. Research at NII focuses on information-gathering techniques and systems for information management. The NII in its push for open access and open science drives the development of academic information infrastructures.
SPARC (US primarily) https://sparcopen.org/	Supporting Organisation	SPARC (Scholarly Publishing and Academic Resources Coalition) is a US based advocacy group allied with SPARC Europe and SPARC Japan.
UK-CORR - United Kingdom Council of Research Repositories (UK) https://uk-corr.org/	Supporting Organisation	The professional organisation for UK open access repository administrators and managers.

Themes

We include here overall and specific themes with examples of best practice, innovation or other noteworthy developments where appropriate.

General observations

The repository landscape internationally is complex. In order to understand it better, we classified the initiatives into three categories:

- Infrastructure - by which we meant an overarching framework that addressed one or more components of repositories.
- Tools - by which we meant technical solutions for specific issues.
- Supporting organisations - by which we meant formal groupings of individuals or institutions with a common purpose.

All of these categories are necessary for a fully functioning repository ecosystem. There are both established and emerging initiatives in all of these categories and the relationship between the various initiatives change regularly as new alliances form and others dissolve. An overall trend from this review is of increasing coordination and cooperation, both regionally and internationally, especially as no one initiative can fulfil all the requirements for

a functioning system. This trend is evident at advocacy, standards, practice and technical levels and it is of note that a recent announcement of a collaboration between Dryad and the California Digital Library[9] seeks to bridge a number of these levels. In this regard the importance of international networks such as COAR becomes apparent. On the one hand, the background work that groups such as this do, may seem just as optional: in fact, their international coordination is essential for a coherent global approach. Lack of international collaboration has led to competing standards in some areas (such as metadata schema), and now the trend is towards increasing alignment - for example of RIOXX and SHARE[10] [11] schemas. It will be important for Australia to be aware of the relevant discussions and ideally to be involved in the decision what the global standard should be. One other interesting recent development is from the UK where the British Library, is piloting a shared repository service for research content built on an open source platform. The Library has appointed open access publisher Ubiquity Press to build the pilot repository.[12]
The role of supporting organizations also becomes important in training, where it is evident that proactive, well designed courses are rare, even if communities of practice exist. This training is one area that would be most appropriate at a national/regional level.

As repositories continue to evolve in response to the changing landscape of scholarly research, there is a need to take a holistic approach to the functions of repositories, which now extend far beyond simply being archives of institutions' content. The diversity of repository content from publications to datasets to software is a feature of today's international repository landscape. Any discussion of repositories has to take a wide-ranging view of their functions. Again, the role of COAR and their next generation project[13] is important here.

Finally, it is obvious that in order for the next phase of repository development to be successful it has to be carefully designed and implemented. It will not happen if individual repositories simply go their own way.

Specific key themes

In the process of this review the following themes emerged repeatedly.

[9] "Dryad Partnering with CDL to Accelerate Data Publishing." *Dryad News and Views* (blog), May 30, 2018. https://blog.datadryad.org/2018/05/30/dryad-partnering-with-cdl-to-accelerate-data-publishing/.

[10] Repository Infrastructure Initiatives Agree to Align Networks - 2 April 2014

http://www.share-research.org/2014/04/repository-infrastructure-initiatives-agree-to-align-networks/

[11] Major Repository Networks Agree to Collaborate on Data Exchange, Technological Development, and Metadata - 16 July 2015

http://www.share-research.org/2015/07/major-repository-networks-agree-to-collaborate-on-data-exchange-technological-development-and-metadata/

[12] https://www.bl.uk/press-releases/2018/july/shared-research-repository-announcement

[13] "COAR Next Generation Repositories: Vision and Objectives." Accessed May 24, 2018. http://ngr.coar-repositories.org/.

Interoperability

The overwhelmingly most important concept is that of interoperability - nationally but even more crucially, internationally. This one concept underpins all recent successful initiatives and is a notable stumbling block when it is not addressed. Notable examples include that lack of standardisation in the interoperability of formats of metadata and in the adoption of FAIR principles - not just for data[14] but for research outputs more generally[15].

Relationship/interdependence/coordination between initiatives

There was a high degree of co-dependence in some areas and this contributed to the success of examples. For example the IRUS-UK project arose out of a specific need identified by Jisc and is dependent on other infrastructure developed by Jisc. When done well this coordination of organisational resources is highly efficient.

Funding

Most of the successful initiatives that we reviewed had addressed the need for stable secure funding, even if they did not themselves have it at that time. Conversely, for a number of initiatives the lack of sufficient stable funding led to a restriction in what could be achieved and an over-reliance on the work of part-time hours limited staff and volunteers - for example COAR. Lack of funding was a specific issue when it came to the retention (or not) of key individuals.

Importance of national/regional networks

Though international networks were important, national and regional ones were crucial especially when they addressed a specific, local need. Good examples of this that we identified are OpenAIRE and La Referencia

There is a constantly changing environment

It was clear that much of the innovation in the examples we identified arose from a response to a changing external environment - driven by changes in scholarly communications and in particular, changes in technology - and a perception that it was important to not be left behind. As a result of this, we note the need for updates of this review on a regular basis.

Recommendations

Based on this review we make the following recommendations.

1. *CAUL endorse the review of international research repository infrastructure and developments.*

2. *CAUL seek specific project funding to develop required national/regional repository infrastructure.*

It is notable that the substantial amount of work on all the current CAUL projects has been done by volunteers, including library practitioners from CAUL members, but also including individuals from outside of CAUL. A rough estimate of the work to simply pull this one report

[14] https://www.ands.org.au/working-with-data/fairdata

[15] https://www.fair-access.net.au/

together is around 40 hours. Developing the required regional repository infrastructure will require ongoing oversight and coordination of a number of diverse activities including the maintenance of regional and international relationships. There is a need to professionalise these activities.

3. CAUL seek a consortial membership of COAR

Given the substantial amount of ongoing work internationally, it is important to ensure we tap into effective networks that already exist - both as a source of advice and in order to influence ongoing international discussions. COAR has clearly emerged as the leading organisation in this area.

4. CAUL set up an ongoing repository technical advisory working group

Building on the work of this CAUL repository project, the next step is a clear articulation of the technical requirements, specifically around interoperability. Specific examples of the work required would be to review and make recommendation on which protocols should be adopted for exchanging repository metadata records between machines e.g. from IR to national harvester like TROVE or to international projects such as OpenAIRE; investigate moving from OAI-PMH to ResourceSync; provide advice and recommendations on how to participate in technical initiatives like Scholix and re3data.org. This group should be supported by a paid individual to act as project officer.

5. That CAUL set up a group which reviews the training and professional development required for repository staff.

The rapidly changing nature of repositories means that repository staff need to develop and maintain new skills. The 2011 survey conducted by Natasha Simons and Joanna Richardson[16] evidenced this need and the authors have repeated their survey in 2017 (results yet to be released). Joining COAR would be beneficial in this regard as they provide repository skills training as a part of membership. CAIRSS could be resurrected as a pro-active ANZ repositories group with the responsibility to ensure sector-wide capacity and capability is developed in order to support the strategic objectives.

[16] Simons, N. & Richardson, J., (2012). New Roles, New Responsibilities: Examining Training Needs of Repository Staff. Journal of Librarianship and Scholarly Communication. 1(2), p.eP1051. DOI: http://doi.org/10.7710/2162-3309.1051

#3: Repository User Stories

Authors

Maude Frances, Associate Director, Library Digital Repositories, UNSW (Lead)

Belinda Tiffen, Director, Library Resources Unit, UTS

Eleanor Colla, Researcher Services Librarian, UNE

Executive Summary

Over the last decade academic and research institutions in Australasia and internationally have developed repository services for managing and disseminating research publications, research data and other digital collections. The need for a review of the Australian repositories landscape has been identified by CAUL as part of its Fair, affordable and open access to knowledge program. As reported in Deliverables #1 and #2 of the review, repository services at many institutions are based on mature technologies, which struggle to keep up with emerging policies and technologies for research infrastructure and scholarly communication.

The user stories summarised in this report document common user scenarios compiled by the CAUL Repositories Project Work Package #3 Working Group. The scenarios were based on input from a range of colleagues in Australian universities and related entities, and by prior work undertaken by groups such as COAR. The user stories are collated according to several stakeholder groups: depositors, university administrators, librarians and repository managers, funding bodies, publishers, and end-users from academic, research, government, industry and community groups. User stories relating to machines or systems are also included. The stories are recorded in a matrix and arranged according to their relevance for Findability, Accessibility, Interoperability and Re-usability of repository content.

This report summarises the user stories, highlighting priorities from the stakeholder groups with reference to FAIR principles.

Findability

Users from diverse stakeholder groups need to efficiently and unambiguously deposit, discover, retrieve, attribute and cite repository content. These outcomes depend on extensive and standardised metadata, which comprehensively incorporate unique identifiers, and which are indexed in generalist and domain-specific browsers and aggregators. Metadata needs to include sufficient contextual information about provenance, rights and research methods, and provide links to data and full-text resources, in order to maintain the integrity of repository content and facilitate ongoing and meaningful discovery.

Accessibility

Stakeholders need to seamlessly retrieve and transfer content and to determine conditions for access. This requires adoption of standard identifiers and internet protocols, which are non-proprietary and WCAG-compliant, and which support preservation of repository content. When content requires mediated or authenticated access, because of data classification and legal requirements, access conditions should be clearly stated in the metadata.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Interoperability

Human and system stakeholders want to aggregate, compare, harmonise, transfer and share content and notifications within and between repositories, computer applications and networks. These processes depend on content and protocols which adopt common standards and shared languages, including machine-readable descriptive and rights metadata. Interoperability requires comprehensive adoption of unique identifiers for people and organisations, related data and publications, research instruments, project and funding information, and classifications such as disciplinary subject headings and FOR codes.

Reusability

Researchers need to be able to reproduce their own and others' research results, and to re-use and combine data from various sources to answer new research questions. Data re-use and re-purposing requires rich and accurate metadata about provenance, rights and research methods, so that all users understand the ethical and legal conditions for secondary use of the data and can replicate the conditions under which the data were generated and analysed. Provenance and version details need to be clearly articulated in the metadata so that future users can unambiguously identify, attribute and cite the data.

Introduction

This report is part of the Fair, affordable and open access to knowledge program established by the Council of Australian University Librarians (CAUL), Review of Repository Infrastructure

The project developed a series of user stories which can inform assessment of current state repositories and development of ideal future research repository infrastructure and services as part of the CAUL Review of Australian Repository Infrastructure. The user stories are intended to represent various stakeholder groups which create and use repository services and content, and the diverse range of research outputs which repositories curate and make accessible.

Method for generating user stories

A matrix was developed to capture user stories relating to nine broad stakeholder groups according to 22 functional areas or headings which were broadly grouped under FAIR principles (Findable, Accessible, Interoperable, Re-usable (See Appendix 5). User stories were derived from numerous sources: adapted from existing resources [COAR, DuraSpace], created by the project team based on experience with repository infrastructure, contributed by library staff from other CAUL Working Groups of the current project. Input and feedback was also sought from Library staff at several Go8 and regional universities, and from members of stakeholder groups at UNSW Sydney, University of Technology Sydney, and the University of New England (institutions of Project Team). Feedback and input on user stories relating to funding bodies was received from the Australian Research Council.

Apart from minor formatting and grammatical corrections, and synthesis of some overlapping items, the user stories were not substantially altered. Although this has resulted in stylistic inconsistencies throughout the matrix, the unedited stories are intended to retain the

perspectives of contributing stakeholders. A matrix of the stories, arranged according to stakeholder groups is at:
https://docs.google.com/spreadsheets/d/1spmsH9H6nDfoPcI_p4bPC4qwl6JSsWn7kniy1OrU0Tc/edit?ts=5b762670#gid=0

User stories based on FAIR principles

There is not a single user of digital repositories for research output. Users have diverse purposes and reasons for engaging with OA and other research material. Future repositories need to be informed by these differences and map them to requirements and system specifications

The present user stories and report differ from the COAR report which defines more specifically the technologies needed to implement next generation repositories. The matrix and report approach FAIR from the user perspective, to answer the question: what will alignment with the FAIR principles mean or achieve for the user(s)?

The summaries below do not capture all user stories - they summarise and identify some priorities relating to each principle. The first paragraph under each principle defines the principle, based on GO FAIR and ANDS FAIR self-assessment tool.

F1 – Metadata are assigned a globally unique and eternally persistent identifier

Definition: Globally unique and persistent identifiers, in the form of internet links, are fundamental for repository infrastructure. Assigned to records, metadata elements and data/files, they remove ambiguity and allow humans and computers to interpret content, over time, with precise meaning.

User priorities: Stakeholders depend on identifiers to accurately and unambiguously discover, retrieve, attribute and cite repository content, which can be subsequently embedded in workflows and output for research, teaching and administrative purposes. For researchers and other repository end-users, unique identifiers support meaningful comparison and harmonisation of metadata and data from disparate sources, such as disciplinary classification schemes, licences, funding programs, journals, people, organisations, and research methods and instruments. Comprehensive application of identifiers enables repositories to be part of global networks of research material and other digital resources and assists university and government administrators with planning and reporting. To enable identifiers to be reliably and efficiently included in metadata and data, deposit processes of repositories need to be linked with sources of the identifiers or be able to extract identifiers directly from repository content.

F2 – Data are described with rich metadata

Definition: Rich and extensive metadata records include descriptive content that facilitates discovery, access and reuse of the data being described. Accurate, rich, standardised and interoperable metadata coupled with flexible tools and user interfaces, enables repository users to search, browse, filter and explore research in ways which support the full spectrum of their needs and behaviours: from precise and quick searching for a known item, to browsing within disciplinary collections, to forming new and unanticipated connections from seemingly disparate research outputs. Metadata based on identifiers which can be read by machines and humans, as defined in F1, are fundamental for repository services and are the building blocks for other FAIR principles relating to access, interoperability and re-use.

User priorities: Users are diverse and they search repositories for a wide variety of reasons. They may be an educator seeking a research paper to download and share with students, or a social scientist searching for previous datasets related to their current research question. They may be a grant funder attempting to understand the breadth of research outputs their funding has generated. They may be a machine harvesting records to create a new subject-based collection of open access content. These users want to search, browse or harvest content in a variety of ways, some of which it is difficult for repository creators and administrators to predict. The importance of F2 is that repository infrastructure is able to accommodate metadata that serves diverse purposes and functions.

F3 – (Meta)data are registered or indexed in a searchable resource

Definition: Unique identifiers and rich metadata descriptions are not sufficient for 'findability' on the internet. Registration with and indexing in generalist or domain-specific searchable resources for research publications and data such as OpenAIRE, Dryad, Zenodo, Figshare, and UnPaywall enhance exposure and discoverability, and increase possibilities for indexing by search engines such as Google and/or Google Scholar.

User priorities: User stories across several stakeholder groups identified the primacy of indexing in and discovery using Google and Google Scholar. To optimise meaningful indexing, repository metadata need to comply with widely adopted generic and disciplinary standards. Communication between repositories and aggregators should ensure that updates are distributed and that usage statistics from numerous dissemination and access points can be transparently aggregated. Comprehensive collections of research outputs, indexed according to provenance, funding source, FOR codes, publisher and other descriptive metadata will benefit government and community stakeholders in tracking resources across multiple institutions, and understanding the commercial and social returns on their investment and collaboration. Optimal discovery of repository items across global web environments depends on systems based on next generation repository technologies (e.g. ResourceSync and Linked Open Data).

F4 - Metadata specify the data identifier

Definition: Datasets should be provided with a unique and persistent identifier, however, as metadata may exist separately from the dataset it describes, it is essential that the metadata includes the dataset's globally unique and persistent identifier.

User priorities: For research owners and parties with an interest in the outputs of research - individual researchers, their institutions, funders or publishers - explicit and unique links between metadata and data enable the immediate discovery of research outputs. Persistent links within the metadata facilitate quick access to full text and open content, from a wide variety of search locations and systems in which metadata are aggregated. Through inclusion of data identifiers within standard metadata schema, wide dissemination can be automated to allow systems to mine, download, transfer and aggregate data, independent of the actual location of data. Persistent connections between data and rich metadata, with sufficient contextual information about research methods, provenance and rights, support ongoing use and re-use of data and provide conditions to enable research findings to be reproduced and verified, potentially optimising the commercial and social return on research investment.

A1: (Meta)data are retrievable by their identifier using a standardised communication protocol

Definition: Ideally users would like to retrieve appropriate internet content directly and unhindered once they have located it. Internet protocols (e.g. http and ftp) define rules and conventions for communication between devices, and tools and services are available to facilitate this process, e.g. APIs.

User priorities: End-users from different stakeholder groups commented on searchability and discoverability being important for accessing data in repositories. Metadata that can be found using standardised persistent identifiers assist discoverability as well as ease-of-use and re-use of repository content. These processes should not depend on specialised tools or communication methods, nor should protocols be limited by components which require manual human intervention, except where these are required for sensitive data or other material that requires mediated access. Protocols should align with standards which allow systems to communicate and integrate with each other to make items seamlessly accessible and should allow integration between systems - for example transferring items from one repository to another. Standardised communication protocols also support user requirements relating to attribution and citation, and for persistent retrieval of repository content.

A1.1: The protocol is open, free and universally implementable

Definition: According to A1.1, for maximisation of data re-use, the protocol should be free (no-cost) and open-sourced, and thus globally implementable to facilitate data retrieval. Anyone with a computer and an internet connection can access at least the metadata.

User priorities: Repository infrastructure in academic institutions is intended to broaden possibilities for dissemination of research output. To cater for various end-users and the range of institutions and organisations they represent, different socio-economic environments in which they conduct business, and varied geographic locations in which they reside means that metadata protocols must be available at no cost and be open-sourced (not proprietary). This will also allow for continuity and a relatively seamless experience for end-users who change repositories (either through moving between institutions or when an institution changes their repository) and will avoid questions about ownership of repository content that is moved or migrated between systems. Protocols must comply with Web Content Accessibility Guidelines (WCAG) so that repository content can be meaningfully accessed and navigated by people with disabilities. Free, open and universally implementable protocols also allow for a wider community conversation about the use and advancement of repositories, and not segregate individuals or institutions according to budget, resources or institution type.

A1.2: The protocol allows for an authentication and authorisation when required

Definition: According to A1.2, the ‘A’ in FAIR does not necessarily mean ‘open’ or ‘free’. Provided that the exact conditions under which data can be accessed are clearly displayed or communicated, even heavily protected and private data can be FAIR.

User priorities: Several user stories from researchers and other stakeholders articulated the need for repository systems and workflows to classify and protect sensitive data. Access to data or metadata can be through an authentication and authorisation process, as required when data are sensitive and access controls are mandated. Ideally, the process will be automatically managed by a machine without requiring human intervention. This may be through a registered username and password associated with an individual or group.

However, the user stories indicate that access to some resources will always require mediation by data owners or stewards. A mechanism for this mode of access may be implementation of a 'request a copy' button, which triggers communication with the party responsible for managing access to the document or data. Systems may need to determine the level of access or embargo period associated with particular items, as well as who has permission to manage and make changes to rights and permissions. Administrative and management stakeholders want functionality which enables authorised people to create, access, and edit metadata on behalf of other people.

A2. Metadata should be accessible even when the data is no longer available

Definition: This principle acknowledges that datasets may degrade or disappear over time because there is a cost to maintaining an online presence for data resources. If resources are not well preserved links may become invalid and data may be inaccessible. Hence, principle A2 states that metadata should persist even when the data are no longer sustained. A2 is related to the registration and indexing issues described in F4.

User priorities: It is important that datasets and other digital assets are well described in metadata records. Metadata provide contextual information about people, institutions, procedures and publications which may be useful if the data or digital asset is not available. Various stakeholder groups emphasised that persistence and preservation of research output and other digital assets is a research and administrative priority for them. Broad dissemination and indexing of repository content with various aggregators, as per F3 above, increases the likelihood that metadata will persist over time. Regular automated checks of links in metadata records will also optimise the ongoing integrity and relevance of the metadata.

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

Definition: This principle stipulates that humans and computers should be able to exchange and interpret each other's data. For computers, interoperability means that each computer system at least has knowledge of data exchange formats that are commonly in use by other systems. This is enabled by well-defined frameworks for describing and structuring metadata and data, and adoption of commonly used ontologies with globally unique and persistent identifiers (F1).

User priorities: Several user stories relating to findability and accessibility (F1, A1) are also applicable to the present (I1) principle in that interoperability of formal and broadly recognised languages for metadata leverages unique identifiers and communication protocols which underpin recognition and retrieval. As in A1.1, end-users expect metadata to comply with WCAG so that repository content can be interpreted by assistive technologies such as screen readers and voice recognition software. Metadata and data which are described using shared and languages can be aggregated, compared and harmonised by researchers, administrators and other stakeholders and can semantically communicate with and move between computer applications and networks. Licences can be assigned and interpreted unambiguously when they are represented using open standards such as Creative Commons, and information about funding sources can be used to link research assets and for administrative and reporting purposes if it is represented in a standardised format. As with F1, deposit processes of repositories need to be linked with authoritative sources so that metadata and representations

are accurate and can be updated in a timely manner as standards and languages evolve. Open standards and widely-adopted metadata are also important for automating the population of repositories from source material, for example, peer-reviewed author manuscripts submitted to journals. An important consideration in a shifting scholarly communications and repository infrastructure environment is migration to new technologies and services - formal metadata that are widely recognised by machines will promote agility in this landscape.

I2. (Meta)data use vocabularies that follow FAIR principles

Definition: FAIR-compliant controlled vocabularies need to be documented and accessible using globally unique and persistent identifiers, which can be discovered and accessed easily by humans and machines. The principle covers universal schema such as Dublin Core or DataCite Metadata Schema, as well as domain-specific vocabularies and standards like MESH or DDI.

User priorities: To accommodate diverse repository users, and their motivations for accessing and engaging with repositories, metadata should be indexed according to recognised and flexible vocabularies. Widely recognised and understood vocabularies and ontologies for conceptualising, describing, classifying and sharing data offer a unified and standardised approach to indexing metadata in a variety of systems and services that are used by stakeholders, such as those required for funding applications, research planning, data analysis and sharing, publication, and research reporting. Machine-readable integration with widely adopted international and regional vocabularies such as MESH or FOR meet requirements among repository end-users for descriptive metadata that can be easily identified and linked across multiple research outputs, activities and administrative documents, and can be discovered and aggregated in global research networks.

I3. (Meta)data include qualified references to other (meta)data

Definition: This principle is about enrichment of data and metadata by cross-referencing or meaningful linking. Drawing on Linked Data technologies, metadata from various web locations can be shared, read, aggregated and interpreted between computers so that researchers and other stakeholders can easily find related resources.

User priorities: There were numerous user stories across all stakeholder groups that relate to this principle. In summary, repositories need to enable dynamic linking to related resources and to information (metadata) that gives context to the people, workflows and content associated with resources. Primary examples for repositories include semantic links between metadata within the repository, between published and OA versions of research output, between publications and associated research data, and links to sites with more information about authors. Meaningful (semantic) links also enable a dataset to be linked with instruments and methods used for data analysis, and, importantly for preservation and enduring access, to the software version associated with a dataset. This may be done by linking to a technical registry such as PRONOM. Identification of versions of resources and mechanisms to ensure that external aggregators and databases track and expose versions of resources are covered under this principle. Other contextual information relates to information about citation of repository works and metrics on research impact (e.g. Altmetrics) and for learning and teaching, links between repository content, reading lists and other course material. Navigation and movement between metadata and resources in various internet locations needs to be

seamlessly experienced by the end user who may want to engage with repository content from other research-related services and workflows (e.g. ethics or data storage sites). For non-academic stakeholders, including government and community research partners, knowledge exchange that is supported by Linked Data technologies adds meaning to research output by giving it context.

R1. Meta(data) are richly described with a plurality of accurate and relevant attributes

Definition: Rich metadata descriptions (as described in F2) enable discovery, however it is important to also capture the purpose and context under which data are generated, such as a description of experimental protocols, or specifications of the tools used to capture data. These types of attributes enable researchers to assess whether data are relevant or useful to them and maximise possibilities for data reuse and reproducibility of research. The data and metadata should retain their initial richness - data publishers should not attempt to predict how, by whom, and for what purpose their data will be re-used, but provide a 'plurality' of attributes that may be relevant for present and future research initiatives.

User priorities: Richly described metadata are a fundamental requirement for data re-use within scientific disciplines, which may require complete and accurate information about the instruments used in data collection and the conditions under which experiments and analysis occurred. In particular, reproduction of research results depends on replication of the conditions under which the original results were derived - these conditions can be specified in study- or variable-level metadata. Similarly, in humanities and social science research, voice recordings, interview transcripts and documentation such as ethnographic notes, require descriptive and technical metadata that contextualises content and format: to understand the equipment or language required to access the data. Detailed contextual information about conditions (who, when, how, why) under which original recordings were made or data were collected will provide important nuances which aid interpretation of the material and its potential re-use. Re-use of survey data, likewise, requires meaningful labels and attributes, along with contextual information which may be documented in code tables. Given the wide range of research methods and metadata schema that are used in different disciplines and research areas, it is not possible for specific requirements of all researchers to be met in a generic repository of research data. A framework in which generic repositories (which focus on discovery and access to content) interoperate with disciplinary archives and data journals (which articulate the relevant attributes and context within disciplinary standards) may maximise possibilities for re-use of the data.

R1.1. (Meta)data are released with a clear and accessible data usage license.

Definition: Principle R1.1 is about legal interoperability - about clearly defining rights of use relating to data. Machine-readable licences summarise the key freedoms and obligations in a format that can be interpreted by software systems, search engines and other technologies.

User priorities: For researchers depositing their research outputs, whether those outputs are in the form of published articles or datasets, a key requirement is the ability to define permissions and the conditions under which material can be accessed, cited and reused. Being able to attach clear and accessible data usage licenses, such as Creative Commons (CC) licenses, in forms that are both human- and machine-readable, allows researchers to deposit their research with the confidence that their work will be used ethically and legally. This is extremely important where work has commercial applications, or cultural sensitivities,

and where data are subject to conditions of data providers or commitments made to research participants. Similarly, for university administrators and grant funders, who may have a stake as data owners or custodians, it is critical they trust repositories to manage access to their data according to legal, ethical and policy requirements. Metadata which includes data usage terms in clear, human- and machine-readable formats are essential requirements for enabling appropriate use of research outputs and other digital assets. Readily available licence terms are equally important for end users, for whom uncertainty about how data and publications can be shared, reused and cited may be a barrier for re-using or re-purposing existing research outputs in their own work. Clearly defined and readily available licences encourage re-use of research outputs, which, in turn, increases the return on research investment and may lead to new insights and discoveries.

R1.2. (meta)data are associated with their provenance.

Definition: Richly descriptive metadata, including a plurality of attributes about the context under which data are generated (as described in F2 and R1) must include metadata that describes the provenance of a research output - clear information about its origin and history. Metadata should include information about the original creator of the data or output, the roles of others who contributed in some way (including to third party content) and subsequent contributors if the data or publication are derived from other sources. Provenance answers questions about why and how the data were produced, where and when and by whom.

User priorities: Researchers and other stakeholders who reuse data need to understand where the data came from, with metadata that enables appropriate attribution and citation. End users, who may be academic researchers, industry or community groups, or educators working with students, need to understand the origins and authenticity of data they plan to re-use or re-purpose, to determine whether the data are credible and trustworthy, to reproduce or validate research results, and to enable appropriate attribution and citation. They need to understand how the data have been processed, and whether they were derived or transformed from other data. In other words, the metadata need to tell a story about where the data came from, where they have been and what happened along the way, with specific information about people, roles and organisations associated with the data or output. To be interpreted by other systems, machine-readable metadata need to use standardised identifiers for people, research methods, instruments, software, file versions, and licences.

R1.3. (Meta)data meet domain-relevant community standards.

Definition: Principle R1.3 relates to disciplinary or domain specific standards. If formal standards or best practices for data archiving and sharing exist in a particular research discipline or field, they should be followed. For instance, many communities have minimal information standards (e.g., MIAME, MIAPE).

User priorities: Researchers can more easily re-use data when the metadata and data are rendered using established formats and standards. Comparison, harmonisation and synthesis of data from various sources will be more efficient and meaningful if data are organised in standardised formats and use common vocabularies, taxonomies and schema.

Recommendation

1. CAUL endorse the report on repository user stories.
2. The user stories inform development of future Australasian repository infrastructure

Related documents

Deliverable #3, User stories by stakeholder group

https://docs.google.com/spreadsheets/d/1spmsH9H6nDfoPcI_p4bPC4qwl6JSsWn7kniy1OrU0Tc/edit?ts=5b762670#gid=0

References

Confederation of Open Access repositories (COAR). (2017). *Next generation repositories: behaviours and technical recommendations of the COAR Next Generation Repositories Working Group*. https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf

Australasian Repository Interoperability Group. (2017). *Repository Interoperability Group progress report - 16 June 2017.* https://docs.google.com/document/d/1UqtVpEQSB84hWavwqMx0grC23G1yZgTEzplnULFQZ88/edit#heading=h.gc7pn0jd40mb

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

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

Authors:

Alexander Sussman, Associate Director, Academic Services, UNSW (Lead)

Margaret Pembroke, Digital Scholarship Librarian (Systems) Southern Cross University

Paula Callan, Scholarly Communications Librarian, Queensland University of Technology

Katya Henry, Research Support Librarian, Queensland University of Technology

PROJECT DESCRIPTION

The Research Repositories Infrastructure project will review the current Australasian open access research output repository infrastructure to provide a current assessment and recommendations for future development to improve: repository interoperability; compliance with institutional, funder and FAIR policy statements; accessibility and the user experience.

The project has been divided into 7 work products.

This report has been written for Product 4: Recommend and make improvements to the current Australasian research repository infrastructure (improve and make the most of what we have).

INTRODUCTION

This report is part of the Fair, Affordable and Open Access to Knowledge Project led by the Council of Australian University Librarians (CAUL). The objective of this report is to provide a list of carefully considered and prioritised improvements that can be made to Australian research repository infrastructure.

To prepare a final prioritised list the product team analysed the reports created by products #1, #2 and #3, sought feedback and input from the product authors, and collated recommendations derived from each report. The collated recommendations (provided as a reference in appendix 6 and listed using FAIR headings) identified common themes that formed the basis for the final five prioritised recommendations.

Analysis of the reports of Products #1, #2 and #3 was undertaken in the following areas of focus:

- Interoperability
- Institutional OA policies
- Metadata
- Compliance with the FAIR principles
- Funder compliance
- Other, as required

REVIEW OF PRODUCTS 1, 2, AND 3

Product #1: Highlights and Analysis

The focus of Product #1 was to review the current institutional repository infrastructure in the Australasian universities sector. The methodology for the review consisted of a survey and analysis of the responses. The findings were broadly categorised as:

• Repository infrastructure and content	• Open access
• Grant funder OA policies	• Research activity identifiers
• Repository interoperability and integration	• Repatriating discipline specific repository records
• Strengths and weaknesses of the repository/infrastructure environment	• Preservation strategy
• Self-Assessment of the FAIR principles	

Table 1. Product #1: Key Quotes and Analysis below comprises direct quotes (selected highlights) from the analysis of Product #1 (in italics) above analysis from the Product 4 Team in relation to the scope of Product #4. Not all of the categories from the product findings are referred to in the table whilst others are included as a different category heading.

Table 1. Product #1: Key Quotes and Analysis.

Review and report on the current Australasian research repository infrastructure Purpose: Provide an account of the current Australasian research repository infrastructure as a starting point for the project.	
Interoperability	*#1: All respondents answered that they were using the Open Access Initiative - Protocol for Metadata harvesting (OAI-PMH) to share metadata records. Sitemaps and Application Program Interfaces (APIs) were the next most common technology used to integrate systems and share data.* **Analysis:** Interoperability is dependent on the ability to identify and ingest records from other systems. The use of metadata and identifiers becomes critical in enabling identification, deduplication and disambiguation of authors and outputs. For these reasons interoperability is identified as the core commonality from all products. It was noted that the survey responses for this criteria span many activities including the ability to ingest and export in different formats.
Metadata	*#1: Rich and standardised metadata were proposed by several institutions as evidence of alignment with findability.*

	Analysis: It is not unexpected that no standardised metadata tags and qualifiers are identified from the responses since such identification is out of scope from the original Product purpose or description. Nevertheless, several institutions still felt the need to identify the correlation between enhanced description with findability (and ultimately reusability). Standardised metadata tags and qualifiers for Australasian university repositories is a recurring theme from Products #2 and #3.
Funder Compliance	*#1: Grant funder policies were supported by 33 institutions… the repository supported the Grant funder policies passively, by virtue of the requirement to support depositing of open access research outputs in general or NHMRC and ARC grants in particular...Monitoring for compliance with the policy was reported positively by 14 institutions...The compliance process was managed by the library in nine answers and by the research office in five...There was no apparent pattern for why they did or did not monitor compliance with grant funder policies.* **Analysis:** The adoption of institutional policies requiring deposit of research output (and now data) to comply with external funder requirements is not universal. This is an area where CAUL can provide guidance and assistance. The aspects around measuring compliance need further discussion. There is scope for a discussion on an agreed standard for measuring compliance, not just in terms of actual repository size but in relation to funded research. Agreed standardised metadata, preferably automated, would assist.
Institutional OA Policies	*#1: Of the 45 respondents, 16 have an institutional Open Access (OA) Policy; 15 have a partial OA Policy, and 14 have no OA Policy. 20 respondents expect change to the current situation within the next 12-24 months: with this spread between institutions which already have an OA Policy (six expect change), a partial OA Policy (nine expect change), or without a policy (five expect change).* **Analysis:** The adoption of institutional Open Access (OA) Policies is significantly less than policies around funder compliance. The responses indicate work on institutional OA policies is currently underway in institutions where it does not yet exist. 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.
FAIR Principles Compliance	*#1: In general, repositories supported the FAIR principles for findability and accessibility more than those for interoperability and reusability...It is important to note that comments about alignment with FAIR principles depended on the respondents' understanding of the principles. It was clear from the responses that levels of understanding varied greatly across institutions.*

	Analysis: Positive self-assessment in support of findability and accessibility ahead of interoperability and reusability is in line with comments and findings from Product #3. Additionally, varied levels of understanding of the application of FAIR principles is also supported by the analysis of the other Products. As understanding the application of FAIR principles is a fundamental requirement there is scope for CAUL guidance on agreed definitions and application. The self-assessment responses for Interoperability would indicate that as the most challenging principle to determine. While that may be the case, it also provides the greatest benefit to repositories, researchers, funders and institutions. Focusing on ways to improve interoperability and providing guidelines/standards around how to achieve that may see a huge benefit to the Australasian repository environment.
Other	**Preservation** #1: *Less than one third (13) of the respondents answered that they had a digital preservation strategy for their research infrastructure and three institutions reported that they are developing a strategy. Six of these institutions either declared, or linked to, a preservation policy document.* **Analysis:** Guaranteeing and securing access and preservation to institutional research outputs requires the adoption and adherence to a preservation strategy. The problem of long term preservation is exacerbated where data are stored in proprietary formats. Technical advice on preservation from CAUL would be beneficial to repository administrators. **Infrastructure Age** **Analysis:** Additional information derived from Product #1 included age of infrastructure though this was not a specific question. Only two identified Infrastructure as new - five as a mix of mature/new - and 36 as mature or older (one declined to comment). The general conclusion is that most infrastructure is mature, older or aged. Significantly, 20 institutions identified they are planning to change infrastructure within the next 24 months. Advice from CAUL on minimum requirements for repository systems (including next generation systems) would be very useful.

Product #2: Highlights and Analysis

The focus of Product #2 was an investigation of the international environment in which Australian research repository infrastructures operate and evolve. This was accomplished

through reviewing initiatives within three broad categories: Infrastructure, Supporting Organisations, and Tools. Within each of the above categories several key examples were identified and then closely inspected in relation to:

• Name	• Description	• Software (open or closed)
• Metadata requirements	• Recommendations for interoperability	• Funder requirements
• Policy requirements	• Key relationships	• Funding
• Compliance with FAIR principles	• Support available	• Other notes

Not all the categories above were applicable to every identified example. Based on the analysis of the key examples, the report identified specific key themes of international developments including:

• Interoperability	• Relationship/interdependence/coordination between initiatives
• Funding	• Importance of national/regional networks
• A constantly changing environment	

As per Table 1, *Table 2. Product #2 Key Quotes and Analysis* below comprises direct quotes from the report above comments in relation to the scope of Product #4.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Table 2. Product #2: Key Quotes and Analysis.

Review and report on the international research repository infrastructure and developments Purpose: Provide an account of international research repository developments which can be used to benchmark Australasian research repository infrastructure.	
Interoperability	*#2: The overwhelmingly most important concept is that of interoperability - nationally but even more crucially, internationally. This one concept underpins all recent successful initiatives and is a notable stumbling block when it is not addressed.* **Analysis:** Products #1 and #3 assert the importance of interoperability however the above quote suggests that successful initiatives are dependent on successful interoperability. La Referencia is an example of a network of Open Access repositories where the success of the initiative was directly related to the ability to integrate into the national node.
Metadata	*#2: Lack of international collaboration has led to competing standards in some areas (such as metadata schema), and now the trend is towards increasing alignment - for example of RIOXX and SHARE schemas. It will be important for Australia to be aware of the relevant discussions and ideally to be involved in the decision what the global standard should be.* **Analysis:** Rich, comprehensive, and standardised metadata using unique identifiers establishes conditions for the possibility for identifying and articulating relationships between researchers, publications, research data, grants, projects, tools, etc. In effect, rich metadata lays the groundwork that can be developed as repositories become increasingly connected and networked. International collaboration is essential for adoption standardised taxonomies. *#2: Notable examples* (of initiatives that have may have stumbled due to lack of interoperability) *include that lack of standardisation in the interoperability of formats of metadata.* **Analysis:** There is an opportunity for a CAUL body (i.e. working group) to review and make recommendation on which protocols should be adopted for exchanging repository metadata records between machines e.g. from Institutional Repositories to a national harvester like TROVE or to international projects such as OpenAIRE.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Funder Compliance	*#2: Most of the successful initiatives that we reviewed had addressed the need for stable secure funding, even if they did not themselves have it at that time. Conversely for a few initiatives the lack of sufficient stable funding led to a restriction in what could be achieved and an over-reliance on the work of part-time hours limited staff and volunteers - for example COAR. Lack of funding was a specific issue when it came to the retention (or not) of key individuals* **Analysis:** An example of the need for ongoing stability of RIOXX became uncertain after funding ceased for many years. In this case, RIOXX appears that it will be incorporated into the OpenAIRE schema. Not all examples of ceased funding have a successful outcome. Conversely, the highly successful infrastructure project OpenAIRE is fully funded by the European Commission.

Product #3: Highlights and Analysis

The focus of Product #3 was to identify and describe a range of personas and use cases for Australasian research repository content and systems which could be used to inform assessment of current state repositories systems and development of ideal future research repository infrastructure and services. This was accomplished through developing a matrix to capture user stories relating to stakeholder groups and functional areas, and then analysing the user stories within the matrix broadly grouped under FAIR principles**.**

Table 3. Product #3: Key Quotes and Analysis.

Review and report on the international research repository infrastructure and developments Purpose: To identify and describe a range of personas and use cases for Australian research repository content and systems, including for research publications (institutional and Research Australia collection), data and NTROs.	
Interoperability	*#3: ...repositories need to enable dynamic linking to related resources and to information (metadata) that gives context to the people, workflows and content associated with resources.* *#3: To accommodate diverse repository users, and their motivations for accessing and engaging with repositories, metadata should be indexed according to recognised and flexible vocabularies.*

	Analysis: A key point from the user stories is that **interoperability** facilitates **findability** as it also streamlines repository processes through the automation of ingestion and discovery of records from other systems. Therefore, user stories related to findability and accessibility were also applicable to interoperability. There were many user stories related to the principle that enriched data and metadata, by cross-referencing or meaningful linking, can aid researchers and other stakeholders to easily find related resources. Additionally, interoperability is critical if repository records are to be discoverable outside the institutional repository, e.g. Google Scholar and Google, and to be reusable for further research.
Metadata	*#3: Globally unique and persistent identifiers, in the form of internet links, are fundamental for repository infrastructure.* *#3: User stories across several stakeholder groups identified the primacy of indexing in and discovery using Google and Google Scholar.* **Analysis:** Users want to search and browse in a variety of ways, some of which it is difficult for repository creators and administrators to predict. Having **accurate, rich, standardised and interoperable metadata** coupled with **flexible user interfaces,** enables uses to search, browse, filter and explore research in ways which support the full spectrum of user needs and behaviours: from precise and quick searching for a known item, to browsing within disciplinary collections, to forming new and unanticipated connections from seemingly disparate research outputs.
Funder Compliance	*#3: For research owners and parties with an interest in the outputs of research - individual researchers, their institutions, **funders** or publishers - explicit and unique links between metadata and data enable the immediate discovery of research outputs*. *#3. ...for university administrators and grant funders, who may have a stake as data owners or custodians, it is critical they trust repositories to manage access to their data according to legal, ethical and policy requirements.*

	Analysis: User stories relating to a funder perspective within a single repository generally related, as anticipated, to compliance. The user comments (from the ARC) about federated searching among multiple repositories to identify research outputs from various funded initiatives once again point to the need for a standardised metadata that facilitates searching (or harvesting). There is evidence for a recommendation for a standardised metadata tag for funding, FOR codes, or other agreed elements to enable comparison of outputs of funded research. Funder compliance could be seen as an institution-specific responsibility and / or as a responsibility of the funder to centrally provision compliance services available to institutions. CAUL could approach major research funding agencies such as the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC). to suggest the benefits of a centralised approach.
Institutional OA Policies	*#3: As a publisher I want to ensure that my Open Access policy is adhered to. (Extract from the User Stories Framework)* *#3: Systems may need to determine the level of access or embargo period associated with particular items.* **Analysis:** Additional user stories (other than the example noted above) related to compliance with institutional policy frameworks or with funder requirements. The January 2018 update of the NHMRC Open Access Policy with its requirement that any output represents a significant peer-reviewed publication arising from NHMRC by made openly accessible within 12 months of publication will make the ability to report on OA compliance within a repository even more important.
Other	*#3: Navigation and movement between metadata and resources in various internet locations needs to be seamlessly experienced by the end user.* **Analysis:** Researchers don't necessarily search repositories through the repository but commonly through a Google-type search or interface. Even if the Google algorithm is not transparent the simplicity of the user-experience makes it an interface of choice for researchers. Simplicity of a repository deposit workflow would also enable depositors to use the system intuitively. Repository UX needs to keep up with other intuitive discovery systems that our users work in, e.g. Google Scholar.

It is noted that assessment of repository user interface was not highlighted by this work. Whereas some importance is placed on user interface and user experience, there is also the

acknowledgement that many users access content directly via search agents such as Google and Google Scholar.

PRIORITISED LIST OF RECOMMENDATIONS

Over 30 recommendations were initially derived from the analysis of Products #1 - #3. These are listed in full in Appendix 6. To generate a reasonable and succinct prioritised list of recommendations the Product #4 team searched for common themes, given that each product report identified many identical issues. These commonalities include concerns about standards (metadata or other, and its consequences for interoperability), the need for authoritative guidance in navigating the Australasian repository infrastructure environment, and general agreement that to make the most of the current repository infrastructure the sector would benefit from a coordinated approach.

General recommendation:

Recommendation 1: It is recommended that CAUL accept and endorse the report on prioritised recommendations to make improvements to the current Australasian research repository infrastructure.

CAUL has capacity to take a leadership role, providing advice and recommendations and with that context the Product #4 Team has collated all the recommendations and advice from reports #1-#3 into 6 priority recommendations:

Proposed Prioritised List of Recommendations:

Recommendation 2: 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).

Recommendation 3: That **CAUL set up an ongoing Repository Advisory Working Group**. Initial objectives for the CAUL repository technical advisory group should include:

Recommendation 4: that the proposed CAUL Repository Advisory Working Group **advise on minimum metadata standards**. The advise will provide consistent terminology and tags to assist in discovery of Open Access material, compliance with government funding requirements, and disambiguation of Authors and Institutions.

- *Rationale:* Reports #1-#3 all noted that granular and consistent metadata would greatly assist with interoperability with import and export of records (and benefit funder compliance). Recommendations on metadata standards and protocols to be adopted should take into consideration international developments in this area as per Product #2 recommendations.

Recommendation 5: that the proposed CAUL Repository Advisory Working Group **develop an Open Access Policy template** for member Institutions with recommendations of what could be included and the wording that should be used to protect against litigation.

- *Rationale:* In the absence of a national Open Access Policy the benefits of an institutional policy include clear guidance for researchers on the institution's stance on Open Access and alignment of the University policy with that of funding bodies such as the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC). Both institutional and funder system centric approaches to ensuring funder compliance for open access could be explored. Further benefits are the ability to reduce reliance on subscription materials and copyright licenses & permissions as more materials are available on open access and as open educational resources.

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.

Recommendation 7: that the proposed CAUL Repository Advisory Working Group provides advise related to **metadata standards and technology for sharing across Open Access platforms and standards.**

Recommendation 8: that the proposed CAUL Repository Advisory Working Group liaise with **ARDC about institutional requirements for training**.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

#5 Develop and report on an ideal state for Australian repository infrastructure and FAIR access to research more generally

Authors

Martin Borchert, University Librarian, UNSW Sydney (Lead)

Andrew Harrison, Research Repository Librarian, Monash University

Belinda Tiffen, Director, Library Resources Unit, UTS

Ginny Barbour, Director, AOASG

Maude Frances, Associate Director, Library Digital Repositories, UNSW

Janet Fletcher, University Librarian, Victoria University, Wellington

Alexander Sussman, Associate Director, Academic Services, UNSW

Natasha Simons, Program Lead, Skills Policy and Resources, ANDS

Purpose

The purpose of this report is to develop and report on an ideal state for Australian repository infrastructure.

About this report

This report on work package #5 of the CAUL review of Australian repository infrastructure pulls information gained from project reports #1 - #4 to develop and report on an ideal state for Australian repository infrastructure. The report takes a high-level strategic and aspirational view at a potential ideal Australian repository infrastructure and applies the FAIR principles, broken down chronologically, suggesting target years for compliance. The statements apply to the Australian repository ecosystem as a whole and it is intended that these can be applied at both the national and institutional level. Comments are not made regarding the repository infrastructure of individual institutions, although some comments are made regarding specific repository software. The intended audience for the report is Director FAIR Access to Knowledge Program (Council of Australian University Librarians, or CAUL), the CAUL Council, and CAIRSS / CAUL repositories Community. See #7 for work undertaken on the Research Australia concept.

The statements were created from a process of environmental scanning and brainstorming of issues by project team members. They are community based in that there have been numerous opportunities for input resulting in many modifications over time.

They are intended to apply going forward and not be applied retrospectively, although institutions could undertake retrospective work if they chose to.

The statements provided are intended to be aspirational and not compliance based as each institution is responsible for its own infrastructure, policy and practice and excellence. The timelines are suggestions. It is not possible to expect each institution to meet each aspiration within specific timelines. Institutions could test how the ideal state statements apply or do not apply at their institutions.

Recommendations

1. *CAUL endorse the report on the ideal state for Australian repository infrastructure and FAIR access to research more generally.*
2. *The statements around the ideal state for Australian repository infrastructure and FAIR access to research more generally be adopted by CAUL.*
3. *The statements of ideal state 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).*

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Findable			
#	*Statement*	*Comment*	*When*
F.1.	*Rich metadata* All records will feature rich metadata and unique identifiers, provenance, as well as the dataset's globally unique and persistent identifier.		2020
F.2.	*Current* All metadata will be made openly available via the relevant repository within three months of publication.		2020
F.3.	*Comprehensively open* All metadata should be made openly available via the relevant repository within three months of publication. All publications, data and NTROs should be openly available via the relevant repository within 12 months of publication. CAUL and the ARC could consider the pros and cons of using the Excellence in Research for Australian (ERA) research assessment activity as a driver for OA by requiring and assessing OA compliance for all Australian university research.		2020-
F.4.	*Harvested by search agents* Australian university open access institutional repositories will be routinely harvested by a range of search agents including Trove (National Library of Australia), "Research Australia", Google, Google Scholar, BASE (Beilfeld Academic Search engine), CORE, and AOIster.	Trove and AOIster cover library collections. Research Australia portal. CORE is indexed by GS. Google for general access.	2020
F.5.	*Retrievable metadata* Metadata are open, free, perpetual and retrievable using a standard communication protocol e.g. SWORD, API. Where data are not free and open, open information is provided on how to negotiate authentication and access methods.		2020
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.		2020-

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, supporting cOALition S.		2020

Accessible			
#	*Statement*	*Comment*	*When*
A.1.	*Comprehensive* All publications, data and NTROs will be openly available via the relevant repository within 12 months of publication. CAUL to discuss with ARC and NHMRC to sign on with European Research Council (ERC) and cOALition S – Plan S.		2020
A.2.	*Machine readable* All research outputs will be machine readable via OCR’d PDF or XML to support large scale data analysis. Protocols comply with Web Content Accessibility Guidelines (WCAG) so that repository content can be meaningfully accessed by assistive technologies.		2020
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. The ARC and NHMRC will contribute automation workflows and systems which assist research institutions to comply.		2020
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		2018

	open publication is optional. Where applicable, metadata and full text of gold OA publications will be deposited into the relevant institutions repository. CAUL will over time and as subscription renewal opportunities come available, investigate the benefits of Publish and Read agreements to increase the percentage of research outputs being made openly available. The metadata and files from gold OA articles will also need to be captured in institutional repositories.		
A.6.	*Affordable Article Processing Charges (APCs)* Australian universities should 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. Australian Universities do not support payment for APCs for hybrid gold Open Access publishing.		2020-
A.7.	*Copyright compliant* Australian universities will achieve full OA while maintaining copyright compliance. Repository records will link to the published version and will also link to the author final manuscript.		2018
A.8.	*Copyright ownership* Australian universities will own the copyright of their research outputs. The pros, cons and implementation of a Scholarly Licensing scheme will be investigated. This will ensure full green OA compliance is possible and legal. This will support cOALition S.		2020

Interoperable			
#	*Statement*	*Comment*	*When*
I.1.	*Interoperable* Australian repositories will be built upon proprietary or open source repository systems which support interoperable standards and data structures.		2020
L.2.	*Links to grant and funding programs* Machine-readable links to grants and funding programs will enable linking between data and		2018

	publications, supporting reporting to funding bodies. The ARC and NHMRC will assist research institutions with reporting of compliance.		
I.3.	*Minimum number of repositories* Australian university libraries will minimise the number of repositories, recognising that publications, data and NTROs can be co-managed, in order to reduce complexity and cost, and increase simplicity, interoperability and the user experience. The Research Australia collection will be inclusive of all research outputs.		2020
I.4.	*ORCID* Every Australian university researcher will have an ORCID ID to facilitate disambiguation and linking.		2019

Reusable			
#	*Statement*	*Comment*	*When*
R.1.	*Machine readable* All research outputs will be machine readable via OCR'd PDF or XML to support large scale data analysis.		2020
R.2.	*Creative Commons licensing* All research outputs will be made openly available using a Creative Commons licence e.g. CC BY.		2019
R.3.	*Well preserved* All Australian university repositories will have a repository preservation strategy and will benefit from preservation workflows and practices.		2020
R.4.	*Downloadable* Full text papers available via Australian University repositories will be downloadable.		2019

Capacity and Capability			
#	*Statement*	*Comment*	*When*
C.1.	*OA advocacy* Australian universities will be active Open Access and Open Science advocates to universities and the public worldwide.		2018

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

C.2.	*Use of repository technologies is well supported* CAUL Repository Community will benefit from a consortium level internationally connected repository technical support group to provide leadership, program management and support.		2019
C.3.	*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.		2019
C.4.	*Internationally connected* CAUL will benefit from international connections and selected members will join the Confederation of Open Access Repositories (COAR) and be an active participant.	Cost of CAUL/CONZUL membership to COAR https://www.coar-repositories.org/about/join/membership-fee-regulations/ For CAUL/CONZUL to join is €28,200 p.a. or A$46,000. For CAUL only is €23,400 or A$38,000. For CONZUL only is €4,800 or NZ$8,500 or A$7,800.	2019-
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* Australian university repository infrastructure will be financially supported. University and government funding options will be sought to fund the CAUL Next Generation Repositories project.	CAUL will investigate Government, ARDC, DVCR and Library funding options.	2020-

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

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.	CAUL Council to decide on priorities. See #6.	2019-2021
C.8.	*Cost effective* CAUL Next Generation Repository Tools will be cost effective, benefiting from commercial, cloud-based tools and a shared, interoperable repository solution.	Funded consortium approach.	2020
C.9.	*Publisher independent* Next generation CAUL repository infrastructure will be publisher independent in order to ensure long term viability and freedom from conflict of interest and commercial pressures.	Applies to CAUL repository infrastructure. Institutions can make their own choices.	2020

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

#6 Next generational repository tools and general requirements

Authors

Martin Borchert, University Librarian, UNSW Sydney (Lead)

Janet Fletcher, University Librarian, Victoria University of Wellington

Elke Dawson, Manager, Open Research, ANU

Allison Brown, Senior Research Publications and Data Librarian, UNE

Purpose

The purpose of this product is twofold:

1. Provide a list of known repository tools. This could form the basis of a future CAUL consortium repository project to go to Request for Information (RFI), develop consortium requirements, and go to Request for Offer (RFO), negotiation, procurement and implementation.
2. Provide a list of general repository system requirements which CAUL could use to test the marketplace.

Scope Notes

Product #1 focused on available cloud based, vendor supported tools, rather than locally hosted solutions. These may be open source or proprietary (commercial). Repository tools supporting both research publication and research data were included. The definition of repository was taken broadly, inclusive of metadata only and file systems. The intent was the create a list of resources supporting the delivery of trusted research services.

Product #2 general repository requirements was added mid-term within the project timeline when it became apparent this was possible.

CAUL members were requested to provide their institutional repository requirements, and these were received from five institutions. The repository requirements document created is an amalgamation of headings and requirements taken from these.

The product team considered it not possible to conduct RFI, RFO, procurement and implementation in collaboration with a broad range of known repository system providers and vendors without dedicated and funded staffing resources.

Pros and Cons of a consortial approach to repository procurement

Pros	**Cons**
Collaborative purchasing power	Increased complexity
Collaborative negotiation power	Meeting diverse requirements
Collaborative expertise	Requires agreement
Common requirements	Agreeing to a basis for cost sharing
Meeting agreed minimum requirements	Integration with existing systems e.g. CRIS
Interoperability	Agreed timing for procurement expenditure
Support community	Implications of members dropping out
Collaboration opportunity	
Raising repository infrastructure to an agreed standard	

Recommendations

1. CAUL endorse the list of repository tools.
2. CAUL endorse the general repository requirements.
3. CAUL consider the pros and cons or a consortial approach to repository procurement.
4. CAUL decide whether to progress a consortial approach to the procurement of a next generation repository system.
5. CAUL develop a project cost sharing basis and fund the project, if applicable.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Repository Tools

Name	Vendor	Cloud based	Locally hosted	Vendor hosted	Open Source	URL	Notes
Open Scholarship Publishing	Ubiquity Press	Yes		Vendor will soon be hosting a general institutional repository system also.	Open tools hosted commercially	https://www.ubiquitypress.com/	Modified OJS with additional services funded by APCs or self-serve. Operates the Ubiquity Partner Network.
Open Conference System (OCS) / Open Journal System (OJS) / Open Monograph Press (OMP)	Public Knowledge Project (PKP)	Yes	PKP offers a full suite of external support options for users of OJS, OCS and OMP who wish to use their own servers.	Can be hosted by PKP or other vendors such a Ubiquity.	Yes	https://pkp.sfu.ca/	PKP is a multi-university initiative developing (free) open source software and conducting research to improve the quality and reach of scholarly publishing.
Dryad	Dryad	Yes	No	Will be	Yes	https://datadryad.org/	Will be offering cloud based institutional instance of the software.

							Non-profit member-based organisation.
Invenio	CERN	Yes	Yes	No	Yes	http://invenio-software.org/	Digital library platform from CERm, featuring research data platform and institutional repository.
Zenodo	CERN	Yes	Yes	No	Yes	https://zenodo.org/	Built on Invenio platform, supports to European OpenAIRE open data policy.
Digital Commons	BePress, Elsevier	Yes	No	Yes	No	https://www.bepress.com/products/digital-commons/	
EPrints	University of Southampton	Yes	Yes	USH will host a simple version of the software for a fee.	Yes	http://www.eprints.org/uk/index.php/eprints-software/	
Figshare	Digital Science	Yes		Yes	No	https://figshare.com/	Initially developed for research data description and sharing, figshare is being further developed by DS for publications and general institutional repository use.
Rosetta	Ex Libris	Yes	Yes	Yes	No	http://www.exlibrisgroup.com/products/rosetta-digital-asset-management-	Digital content management system and preservation system.

						and-preservation/	
Esploro	Ex Libris	yes	No	Yes	No	https://www.exlibrisgroup.com/products/esploro-research-services-platform/	Website states Esploro “moves beyond the standard institutional repository” - works best if you are an ExL library?
Equella	Pearson	Yes	Yes	Yes	No	https://www.equella.net/	Equalla can also be cloud hosted by the vendor. Initially designed as a learning object repository, although now used also for institutional repository purposes.
DSpace	DSpace http://www.dspace.org	Yes	Yes	A managed DSpace repository service offered by the DuraSpace not-for-profit organization- DSpaceDirect is the hosted repository solution for low-cost discovery, access, archiving, and preservation.	Yes	DSpace open source software is a turnkey repository application used by more than 1000+ organizations and institutions worldwide to provide durable access to digital resources.	

				http://dspacedirect.org/			
DSpace CRIS	DuraSpace	Yes	Yes	Yes, there are vendors available to host this platform, including within Australia.	Yes	https://wiki.duraspace.org/display/DSPACECRIS/DSpace-CRIS+Home	This Current Research Information System is a DSpace extension for Research Data and Information Management. Project originally funded by Hong Kong University.
Samvera formerly known as Hyku / Hydra-in-a-Box	Partnership of US universities		Yes	DuraSpace plans to offer a hosted Hyku service named HykuDirect in partnership with a network of service providers. https://wiki.duraspace.org/display/hyku/Hyku+Documentation	Yes	http://samvera.org	Hydra-in-a-Box is a community-led project to extend the existing Hydra project codebase and its vibrant community to build, bundle, and promote a feature-rich, robust, next-generation digital repository that is easy to install, configure, and maintain.
Islandora	Islandora was originally developed by		Yes		Yes	https://islandora.ca/	Open-source software framework designed to help institutions and organizations and their audiences

	the University of Prince Edward Island's Robertson Library.						collaboratively manage, and discover digital assets using a best-practices framework.
CKAN	Digital Curation Centre (DCC) (UK)		Yes			http://www.dcc.ac.uk/resources/external/ckan	Research data repository for metadata and small datasets.
DataVerse	Institute for Quantitative Social Science (IQSS) and Harvard University.		Yes			https://dataverse.org/	Research data repository software.
Fedora	Fedora Leadership Group, DuraSpace		Yes		Yes	https://fedorarepository.org/	Institutional repository software platform.
Sharestream			Yes			https://www.sharestream.com/	Online video platform.
Kattura			Yes			https://corp.kaltura.com/	Online video platform.

PURE	Elsevier		Yes			https://www.elsevier.com/solutions/pure	Pure aggregates your organization's research information from numerous internal and external sources and ensures the data that drives your strategic decisions is trusted, comprehensive and accessible in real time.
Elements	Symplectic		Yes			https://symplectic.co.uk/products/elements/	Capture, analyse and showcase research with the world's leading Research Information Management System. The relationship with Dimensions and Figshare needs to be explored.
Converis	Clarivate		Yes			https://clarivate.com/products/converis/	Assemble the complete professional profiles for a complete and up-to-date collection of all teaching, research, and service related activities. Providing you an overview of all accomplishments, with advanced analytic reports of outputs and impact.

General Repository Requirements

The requirements found in the following tabs have been broken down into five areas:

The Key tab:

Contains priority classifications and a glossary of acronyms used within the spreadsheet.

Sub Headings: Priority, Glossary.

The Non-functional requirements tab: See Appendix 7

Contains UNSW requirements that do not relate to how the repository may function and contains conditions of use.

Sub Headings: Service Provision and performance, Exit Strategy, Content ownership, Compliance, Data Centre location, Data Centre security protocols, Data Integrity and Security, Support, UNSW Branding, Preservation.

The End User experience tab:

Contains UNSW requirements for the public aspect of the repository.

Sub Headings: User Interface (UI), Search, Accessibility, Content engagement, Content access, Interoperability, Metrics.

The Repository management tab:

Contains UNSW requirements for those who administer the repository.

Sub Headings: Administrative User interface (UI), Collection development (ingest/deposit), Content management, Content control, Metadata management, Rights Management, End User access management, Administrative User interface access management, Content delivery, Preservation.

The Functional System requirements tab:

Contains UNSW requirements as to the more technical aspects of the solution functions.

Sub Headings: Reporting, Notifications, Capabilities, Interoperability, Discoverability, Migration

Key Tab

Priorities:

All the requirements can be sorted into one of the following classifications:

Classification:	Meaning:
Core	Must have (deal breaker)
Preferable	Would like to have and would not like to live without
Worthwhile	Would like to have but could live without
Not Applicable	This requirement does not apply to my need

Glossary

Term	Meaning
AMQP	Advanced Message Queuing Protocol
BASE	Search engines especially for academic web resources
BE	Back End (solution administrative staff access)
CHOR-AUS	Support discovery of OA content and assists compliance with funder requirements by working with publishers, funders, institutions and authors
COBIT	Control Objectives for Information and Related Technologies
CORE	Repository for open access research papers
CSA	Cloud Security Alliance
Digital item	All outputs included into the repository for discovery, access and management. The 'content' aspect of the repository (excluding metadata); e.g. text, video, audio, image file formats and computer code formats
FE	Front End
FOR code	Fields Of Research code
HuNI	Humanities networked infrastructure
IDS	Intrusion Detection Systems
IGSN	International Geo Sample Number
IIIF	International Image Interoperability Framework
IPFS	InterPlanetary File System. A peer-to-peer hypermedia protocol
IPS	Intrusion Protection Systems
ISAE	International Standards for Assurance Engagements
ISMS	Information Security Management Standards
JSON	JavaScript Object Notation
METS	Metadata Encoding & Transmission Standard
NIST	National Institute of Standards and Technology
OAI-ORE	Open Archives Initiative Object Exchange and Reuse
OAIS	Open Archival Information System
OCR	Optical Character Recognition
ODRL	Open Digital Rights Language

Glossary

Term	Meaning
RAMP	Repository Analytics and Metrics Portal
RAMS	Records & Archives Management System
SLA	Service Level Agreement
STAR	Security Trust and Assurance Registry
SUSHI	Standardized Usage Statistics Harvesting Initiative
SWORD	Simple Web-service Offering Repository Deposit
ToS	Terms of Service
Trove	Discovery service for Australian resources
UI	User Interface
W3C	World Wide Web Consortium
WCAG	Web Content Accessibility Guidelines

#7 Review and Report on the Potential for a Research Australia Collection

Authors:

Belinda Tiffen, Director, Library Resources Unit, UTS

Hero Macdonald, Director, Learning Services and Digital Innovation, UNSW

Kay Steel, Manager, Research and Strategic Projects, Federation University Australia

Julia Hickie, Assistant Director, Trove Data, Discovery and Delivery, NLA

Catherine Brady, Collection Development Librarian, Australian Research Data Commons

Executive Summary

The Council of Australian University Libraries (CAUL) is currently engaged in a strategic project to review repository infrastructure in Australia and to consider how we can improve the current state to ensure Australian repositories deliver fair, affordable and open access to Australian research as part of the CAUL strategic commitment to the FAIR principles.

As part of this project a working group was established to assess the desirability and feasibility of a "Research Australia" collection. That is, a public portal enabling wide discovery of, and access to, Australian research outputs including publications, datasets and non-traditional research outputs (NTROs).

This particular work package was undertaken in two parts. The first stage aimed to establish the desirability or business case for such a service or collection through interviews with stakeholders from the repository community. The second stage assessed the feasibility of such a service or collection within the current technical, policy and resource environment. Key findings from the review established:

- While there is some interest in a Research Australia Collection, stakeholders failed to identify a compelling case for the benefits of such a service
- Technically such a service would be feasible, but would require substantial and ongoing resourcing to establish and maintain
- The National Library of Australia is currently undertaking a review and upgrade of the TROVE service, and this should be explored as a lower-cost option to provide key functionality of a "Research Australia" style service leveraging established infrastructure.

Recommendations to CAUL are provided at the conclusion of the report.

Since this work was undertaken, the concept of a PubMed Central (PMC) Australia infrastructure and collection for health and medical outputs has arisen. It was not possible to include this in the project report to any degree and it is simply noted here. CAUL could work with stakeholders on this concept.

Stage 1:

Establishing Desirability of a Research Australia Collection

To understand possible interest in such a project, and to explore questions of technical and policy feasibility, in September 2018 the working group conducted 14 interviews with key stakeholders in the Australian repository environment. This was accompanied by semi-formal feedback from focus groups and researcher interviews.

Methodology

Semi-structured interviews were conducted with representatives from academic libraries and grant funding bodies, including University Librarians, directors and repository managers. Interviewees were selected to provide a representative sample from all Australian states, and from each of the identified university groupings (e.g. Go8, ATN, IRU).

A semi-structured interview method was employed, which encouraged exploring unanticipated ideas, but which made codifying answers difficult. In the summary results

below, the number indicates how many respondents expressed an opinion, but due to the methodology these are indicative only.

At the CAUL Research Repositories Community Days held in Melbourne on Monday 29th and Tuesday 30th October 2018, further feedback from the repository community was sought. The project and findings to date were presented in a brief formal presentation, and a subsequent 'unconference' session for group discussion was used to explore topics related to shared collections, services and infrastructure. High-level responses from this event have been incorporated into the Key Findings below.

Members of the working group also informally discussed the concept of a Research Australia service with researchers at their institutions. This represented a very small sampling, and while this feedback therefore helped to reinforce findings from stakeholder interviews, was viewed with some caution as a data source.

Key Findings

- Majority of respondents identified some value in having an easily accessible 'picture' of Australian research outputs
- 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, NTROs, outputs from agencies beyond CAUL institutions)
- 4 respondents questioned scope in terms of defining "Australian" research
- No consensus emerged on mechanisms for delivering such a service
- The majority of stakeholders emphasised the importance of Research Australia making research outputs openly available, but also noted that challenges involved in making research openly available via existing repositories.

Detailed Findings

Benefits

Most respondents could identify some benefits to a 'Research Australia' style service. Some of the key benefits that were mentioned:

- Value for government, funders, universities to showcase national and institutional outputs [5] – with some comments noting research outputs can get lost in TROVE:

  "We all have repositories, and they are being harvested by TROVE, but who knows about TROVE? Only humanities – STEM don't'

  "Comprehensive and reliable discovery and access to Australian research outputs"

- Value for major grant funders to access a comprehensive list of funded research outputs [4]
- Value for researchers in finding collaborators (interestingly the theme of researchers more easily discovering research per se was not raised by most respondents)
- Value for public in having access to a wider range of research (although specific case studies were generally not mentioned) [3]
- For universities and specifically repository managers, better statistics with the potential for national benchmarking; and possibly efficiencies in system, metadata and/or infrastructure management. The latter was dependent on how they imagined a Research Australia collection/portal functioning [2]

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Feedback from the CAUL Repositories Community Day also suggested that a single collection of Australian research would deliver some value to government, funders and potentially the general public, as a way of providing a showcase/single access point to research outputs, however no compelling use cases were articulated.

Increased exposure of Australian research outputs in Google and Google Scholar could also be a benefit of a Research Australia portal. While it was not raised specifically in stakeholder interviews, the findings of Work Package #1: Review and Report on Current Australasian Institutional Research Repository Infrastructure, revealed heavy use of Google, Google Scholar and Trove to index institutional repository data, however noted not all repositories are harvested by these key discovery portals. The report noted: "The high number of Australian contributors to TROVE, nearly on par with the obligatory Google indexing, is evidence of strong interest in having a national portal for the institutional repositories' combined research outputs" (see p. 16 of this report). A desirable outcome to support the FAIR distribution of and access to Australian research would be universal coverage of Australian institutional repositories in Google, Google Scholar and Trove.

Barriers to a Research Australia Collection

Both as unprompted comments and in response to questions, respondents did raise issues with both the need for such a collection/portal, and/or the practicality:

- Several respondents questioned the value of "another portal", with several raising Google and Google Scholar as existing services popular with clients which did not require duplication. [5]

  "Aggregators are disappearing as the need they once met is filled by alternative services and connections. The case for discovery and access is weak."

  "If we are talking about a shared portal, then I don't see a need for that from either a discovery or preservation perspective"

  "Should we just focus on Google, or where people are?" "Why not just use Google Scholar, etc?"

- The necessity of having access to full text content was also raised by many respondents, with some noting that without the full text the value of such a service was limited; although some others did suggest a metadata only catalogue was still useful. [3]
- Other concerns raised included the quality of metadata, with several respondents noting any service would be limited by the quality of its input metadata and lack of interoperability between existing systems [3]
- Several respondents noted that researchers tend to identify more strongly with discipline-based platforms, rather than nationally-based platforms.

At the CAUL Repositories Community Day, there was an overall sense that discovery of research outputs was not seen as the highest priority to focus on. Standards, interoperability and replacing aging repository infrastructure were raised as more pressing needs. Several people also raised the complex internal environment at their institutions as a potential barrier to any shared or collaborative systems - e.g. infrastructure decisions are not made by the Library, integration with internal CRIS systems is priority over external-facing systems.

Scope and Content

Respondents were unanimous that any Research Australia collection should include traditional research outputs from CAUL institutions (although one respondent did suggest it could start with only ARC and NHMRC funded research). Many were in favour of also including grey literature [9 – although one respondent specifically spoke against this] and NTROs [5]. Views on research data, theses, digital collections beyond research institutions and other materials were more diverse.

Statistics

Many saw the ability to get solid metrics, with the possibility of benchmarking, as a potential benefit of a Research Australia collection. There was interest in alternative metrics [4], discipline specific or FoR code metrics and information around OA and embargo periods.

Business Model/Technical Infrastructure

This was probably the area of greatest uncertainty. The interview questions did not posit a particular model for how a Research Australia collection would function. Many referenced Trove in their answer, generally favourably as an existing infrastructure that could be adapted. The option of leveraging existing systems and focusing on better interoperability and linked data, without specific reference to Trove, was also a theme that came through in several interviews. There were, however, a few voices against Trove and a small amount of (unprompted) interest in looking at new and/or shared repository 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."

Concerns were also raised at the CAUL Repositories Community Day about the potential cost of any such service, and the political will to gain the necessary funding and support.

Stage 1 - Desirability Study Conclusions

Views across the sector were generally somewhat mixed on the need for, or desirability of, a Research Australia portal. In considering responses, however, it should be borne in mind that interviewees were not presented with a detailed proposal of how such a system would work, but were allowed to consider how they would conceive such a system. Overall, while views were mixed, it was felt there was enough interest displayed to warrant proceeding to Phase 2 of the work package to consider the feasibility of such a service.

Stage 2:

"Research Australia" Feasibility Study

From interviews with stakeholders conducted during Work Package #7 and review of findings from other work packages, especially #3 User Stories and #5 Ideal State, a list of requirements to create a minimal viable product for a "Research Australia" collection or service was created. These are based on the FAIR principles. Please see Appendix A.

Infrastructure and Business Model

Establishing a completely new 'Research Australia' portal or service would require significant investment in infrastructure development to deliver the full potential of such a collection. Examples from overseas such as OpenAIRE and PubMed and European PubMed suggest

success is only achieved with significant and ongoing funding and where multiple partners are contributing to supporting the enterprise.

Three potential models for how such a service could be delivered have been identified and considered as part of this current study, organised in order of scale of investment required.

1. Upgrade TROVE

Use the existing Libraries Australia infrastructure to establish a 'Research Australia' collection. This would require feature enhancements to achieve the required functionality outlined in Appendix A. Agreement with, and support from, the National Library of Australia would be necessary to achieve this outcome. It would, however, have the advantage of leveraging existing infrastructure and a known and trusted 'brand' in Trove and NLA. The group engaged with the National Library around the project it is currently scoping to make improvements to TROVE, however, the project is not resourced to make the extensive changes necessary to meet the requirements for a Research Australia collection system as outlined above. Cooperation between CAUL and the NLA would be necessary to agree to the scope of such a project and shared resourcing.

In relation to the wider discoverability of and access to Australian research outputs in Trove, NLA representatives confirmed Trove is currently crawled by Google, with Trove items generally ranking highly in Google search results. Trove content is not directly harvested by Google Scholar. This includes the subset of Australian Institutional Repository content. To date, seeding Google Scholar has not been identified as desired functionality for Trove. If a Research Australia portal were to identify this as required functionality, it could be investigated as an enhancement to Trove.

2. Develop a New Portal

Using international models such as OpenAIRE or European PMC as a guide, a Research Australia service could be developed using new infrastructure to harvest data from existing institutional and other repositories into a shared, new portal. This portal could be developed as a custom technical solution, or seek partnerships to use existing infrastructure. Such a model could use software and services such as F1000, which is used by other institutions and funders to power portals to research outputs. To investigate this model the group spoke to Lisa Kruesi who is completing a PhD into the feasibility of an Australian PMC, and sought to understand the business models and user demographics behind similar national and international services. It is noted where these systems have been successful - such as PubMed Central or OpenAIRE - there are large collaborative partnerships behind them, and plans to sustain funding and support. Even using existing infrastructure such as F1000 or similar, developing such a portal would require significant resources. The group is concerned that without having established a compelling business case for a Research Australia portal, the political will and partner network necessary to make such a portal successful and sustainable does not exist.

3. Shared Infrastructure

In the course of interviewing stakeholders, several raised the issue of establishing shared infrastructure, given the Australian context where most academic institutions have repository infrastructure which is at least ten years old. As one interviewee said: "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". Such a model would have the potential to leverage shared purchasing and bargaining power across Australian institutions, and to remove barriers to interoperability by establishing a shared ecosystem. Against these potential benefits, the costs of establishing such an environment,

and the barriers to building shared infrastructure must be acknowledged. As many stakeholders raised in discussing the potential for a Research Australia portal, repository infrastructure and management of both research outputs and technical architecture is a shared responsibility within their institution, and buy-in would be needed beyond libraries to establish such a shared environment. The Working Group felt establishing such an agreement would be possible, but would require significant investment of time. Discussions between CAUL and CAUDIT, as the peak body for technical infrastructure in the Australian higher education sector, could be an initial starting point for such an agreement.

4.Ebnsure indexing by search agents and metadata harvesters

Indexing by search agents such as Google and Google Scholar can greatly increase the findability of research outputs, as can metadata harvesting by specialised research metadata harvesters such as CORE, SHARE and OpenAIRE.

Each institution could or should arrange to have their repository harvested by these agents to increase the findability of their research outputs. The agents would generally combine Australasian research outputs with those of other regions and a Research Australia Collection is not provided as such. Indexing by search agents and meta-repositories could be centralised by a Research Australia infrastructure.

Existing Australian Repository Environment

In considering the options outlined above, it is worth briefly outlining the current repository environment in Australia, and acknowledging it presents significant barriers to each of the above potential solutions.

Currently the Australian landscape is highly diversified, with several different repository softwares in use within the academic sector. The environment is becoming more fragmented as new software is added to support data management and curation (e.g. Figshare, Omeka). The Elsevier product PURE is also gaining some market traction and it is noted this system is not interoperable, raising significant barriers to achieving a Research Australia collection which would meet the FAIR principles as detailed above.

Within this environment, the group noted that any of the above outlined models would require not just centralised resources to set-up such a solution, but funding for institutions to support business readiness. This would need to be focussed on activities such as metadata clean-up, implementation of standards and protocols, and staff training. From the current experience of TROVE in supplying aggregator services, it is estimated that between 25-50% of Australian institutions would need to undertake some level of uplift to be ready for an aggregated Research Australia solution.

The role of the major national research funders, ARC and NHMRC, as well as existing critical, government funded infrastructure and research support services such as Research Data Australia and the Australian Research Data Commons also needs to be considered. These stakeholders have quite specific requirements and in some cases their own strategic roadmaps. Extensive collaboration would be required to ensure alignment of goals and a shared vision across these bodies.

Recommendations

The Working Group conducted interviews and focus groups with stakeholders from Australian academic libraries, research funders and other aligned institutions. It also attempted to undertake some desktop research regarding other national and international efforts in the repository landscape. Through this work it established there is some interest in the idea of a Research Australia style collection or portal, but not clear articulation of the benefits of such a product. While most stakeholders spoken to could identify potential users of such a system, no compelling business case could be established.

The Working Group also considered the reports of other work packages from the CAUL Review of Australian Repository Infrastructure Project to articulate the minimum required features for such a Research Australia collection, and to identify potential models to achieve such a service. In its research the group noted those portals or services which have achieved success and remained sustainable beyond implementation, were well funded and supported by established partnerships of stakeholders. The Working Group is concerned that the desirability study did not provide sufficient evidence of the political will or support to establish and sustain such a network.

Given the Working Group did not establish through its desirability study a compelling need or use case for a Research Australia portal proportional to the resources required to implement and sustain such a service, it therefore makes the following recommendations:

1. CAUL Executive explore partnership opportunities with the National Library of Australia to provide input into the current TROVE enhancements project and incorporate as many of the required features as outlined in this report as possible within the TROVE service.
2. A new standing Repositories Technical Advisory Working Group (Recommendation #5 in the CAUL Review of Australian Repository Infrastructure Report 2018) is established to provide an appropriate liaison between CAUL and NLA to contribute to the current TROVE enhancements project.
3. CAUL work with TROVE to investigate Google Scholar harvesting as a priority for Research Australia.
4. The CAUL Repositories Technical Advisory Working Group prioritises efforts to establish sector wide standards in metadata schema, protocols and language to ensure interoperability of systems.
5. The CAUL Repositories Technical Advisory Working Group prioritises national adoption and use of standard unique identifiers, particularly ORCID, RAID and DOI to assist with implementing the F.A.I.R principles and enhanced reporting.
6. CAUL actively follows the progress of the Australasian PMC project as a potential proof-of -concept for a broader national collection.
7. CAUL conducts a comprehensive consultation process to gauge demand for shared infrastructure and potentially consider funding a project, using requirements from Work Package #5, to go to market to procure a next gen repository system on behalf of its members, and that institutional involvement be “opt-in”.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Appendix A

Required Features

Findable

Feature	Description
Faceted search options on key metadata	Ability to filter search results on: ● Institution ● Funding source ● Access conditions – (open/close/embargo) ● Licence conditions – reuse and data mining ● Unique ID (e.g. RAID) ● Discipline (e.g. FoR)
Linking data and publications	The system should incorporate tags/identifiers or other features to enable related publications and data files to be easily identified by end users.
Harvestable	Records are able to be harvested by widely-used discovery systems including Google and Google Scholar, using internationally accepted standards and protocols.
Collaboration mapping	End users should be able to query the system to see links at national and international levels between researchers (e.g. co-authorship). Ideally this should be able to be visualised and exported in custom reports.

Accessible

Feature	Description
WCAG compliant	Conforms to international accessibility standards.
Branding	The collections of each institution can be branded (including logos, visual identities) to distinguish the collections of individual institutions within the wider 'Research Australia' collection.
Licence conditions related to access are clearly displayed	End users can easily identify access conditions (e.g. open, closed) in a record. This metadata should be both machine and human readable to facilitate both access and interoperability.

Interoperable

Feature	Description
Unique Identifiers	Unique identifiers from standard schema applied consistently (DOI, ORCID, RAID).

Machine-actionable metadata	Metadata is both human and machine readable to support current and future states of interoperation and automation.
Standards based	Metadata schema, standards, language and protocols are agreed, applied consistently and implemented universally.
Exportable in standard formats	Search results should be able to be exported in commonly used formats and to widely-used bibliographic management software.

Re-usable

Feature	Description
Use rights	Use rights/licence conditions are included in metadata in both human and machine readable forms, and records can be filtered on this information.
Provenance	Metadata must be rich enough to establish the source of research outputs (publications and data).

Reporting

Feature	Description
Usage reports	Ability to generate reports by institution showing: ● Access/views ● File downloads
Benchmarking	Ability to generate reports showing comparative data across institutions, including: ● Access/views ● File downloads
Open Access monitoring	Ability to generate reports on the percentage of materials which are open access; reports segmented by licence type.
Exemption reporting	For repository managers the ability to run reports/receive alerts on records which do not adhere to system rules (e.g. missing files, missing required metadata).
Open statistics	Statistics can be openly available at both the individual item level (downloads, alternative metrics, citations) and institution level via a publicly accessible interface.

Appendix 1: Project Initiation Document

CAUL Program:	**FAIR, AFFORDABLE AND OPEN ACCESS TO KNOWLEDGE**
Title:	Review of Australian Repository Infrastructure Project
Document Author (name, email, phone and position):	Jill Benn, CAUL Program Director – Fair, affordable and open access to knowledge and University Librarian, UWA jill.benn@uwa.edu.au / 08 6488 2355
Objective:	The objective of this project is to determine how improvements to repository infrastructure can be made across the sector to increase the findability, accessibility, interoperability and re-usability (FAIR) of Australian-funded research outputs.
Brief Project Description:	Through a review of current repository infrastructure and international developments, this project will identify an ideal future state for institutional repositories, determine what enhancements can be made to improve this infrastructure across the Australian higher education sector and implement changes which enable international discoverability and interoperability of Australian-funded research.
Context:	As outlined in the CAUL paper on Open Access, from 2007 to 2009 Australian universities were provided with funding to establish digital repositories for research outputs through an Australian Government investment of $25.5M under the Australian Scheme for Higher Education Repositories (ASHER) primarily to support the RQF and ERA. These repositories now contain over 1 million items, including Open Access (OA) versions of journal articles, theses and other university publications with more than half freely available for download (though most of these do not have associated OA licences, which specify their reuse rights, nor other consistent metadata). These items were used 36.5 million times in 2016. Many of these research outputs are Author Accepted Manuscript (AAM) versions of published journal articles. Items in institutional repositories are discoverable via Google and NLA Trove. Repositories are an essential piece of research infrastructure, not only aiding the discoverability of research outputs globally but also delivering content for the Excellence in Research for Australia (ERA) assessment exercise. The ARC and NHMRC open access policies require that metadata for research outputs funded by them, and in some cases an OA version of the paper, must be made available to the public in a repository as soon as possible. Repositories have the potential to reduce Article Processing Charges (APC) costs through the deposit of AAMs (green) for OA. However, at present these repositories cannot be easily and accurately harvested or networked due to a lack of metadata and other interoperability standards. There is a perceived lack of visibility of institutional repositories, especially internationally, despite their important function. Some current infrastructure is older and not user friendly and requires more staff input to run. There are a number of options to improve repositories. These range on a spectrum from simply agreeing on common metadata, to shared infrastructure

	to a completely integrated network, and there are a number of international examples which can be considered as part of identifying future state scenarios. In 2016 the Australasian Repository Working Group was established under the auspices of AOASG and the CAUL Research Advisory Committee. This group, co-chaired by Ginny Barbour (AOASG) and Natasha Simons (ANDS) has undertaken a significant amount of work already in relation to interoperability and metadata standards which will be of great benefit to this project. At the CAUL Meeting in September 2017, Members indicated a preference for Green OA and repositories play a critical role in being able to deliver this preference. This was supported at the Universities Australia Deputy Vice-Chancellors meeting in Hobart in October 2017.
Project Owner (name, email, phone and position):	Jill Benn, CAUL Program Director – Fair, affordable and open access to knowledge and University Librarian, UWA jill.benn@uwa.edu.au / 08 6488 2355
Project Lead (name, email, phone and position):	TBC - CAUL office calls for EOI from CAUL members and their delegates
Project Team (name, email, phone and position):	TBC - CAUL office calls for EOI from CAUL members and their delegates
Project Governance Group:	The CAUL Executive will provide oversight and governance in relation to this Project.
Benefits:	An improved repository infrastructure environment will: • Increase the visibility, discoverability and reusability of Australian-funded research outputs internationally and for the benefit of industry and the broader community • Align Australian efforts with international best practices • Increase interoperability between repositories and create efficiencies when Australian researchers move between institutions • Provide more efficient and effective infrastructure across the sector
Scope and Deliverables:	1. Call for Project Lead and Team Members to form a project team to review repository infrastructure, identify an ideal future state including a review of international initiatives such as OpenAire and create a prioritised list of enhancements for findability, accessibility, interoperability and reusability. 2. Seek feedback from CAUL members and other stakeholders. 3. Lead the implementation of changes as required.
Resources & Key Stakeholders:	Stakeholder engagement is critical to the success of the Project. Engagement, communication or collaboration is expected for (but not limited to): 1. CAUL Members 2. Australasian Open Access Strategy Group 3. Australian National Data Service 4. FAIR Implementation Steering Group 5. Universities Australia

	6. Australasian Research Management Society 7. Researchers 8. ARC and NHMRC 9. NLA Resources required: 1. CAUL personnel to lead and participate in this group 2. CAUL office personnel for project management 3. Resources required for implementation of recommendations as identified in the project
Key risks and mitigations:	Risk: Inadequate resources will be available for this project Mitigation: 1. CAUL members will be encouraged to nominate themselves and their staff to participate in this project 2. CAUL members will be kept informed on progress and potential impacts to their repository environment 3. CAUL Office resources will be prioritised to enable appropriate levels of support
Related Projects/ Initiatives:	Other CAUL-related projects under the Fair, affordable and open access to knowledge program, including the Statement on Open Scholarship Review and the Retaining Rights to Research Publications Projects.
Timelines:	Commence January 2018
Attachments submitted:	CAUL paper on Open Access Australasian Repository Working Group Report to AOASG and CRAC, February 2017 Repository Interoperability Working Group Progress Report – 16 June 2017 Example of Schema for Metadata vocabulary - compiled by Australasian Repository Interoperability Working Group, 2017 Repository Network Characteristics - compiled by Australasian Repository Interoperability Working Group, 2017
Document Version:	Drafted Jill Benn 17/11/17 Edits/comments Ginny Barbour, Natasha Simons 21/11/17, Diane Costello 23/11/17, Jill Benn 11/12/17 *Insert document URL here* *http://www.caul.edu.au/content/upload/files/FairAccess/fair2017repository-infrastructure-initiation.docx*

Approval from CAUL Program Director and CAUL President:

CAUL Program Director:	**Signature:**	**Date:**
CAUL President:	**Signature:**	**Date:**

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Appendix 2: Project Plan

Project Lead	Martin Borchert - *University Librarian, UNSW Sydney*
Project Team Members	Andrew Harrison - *Research Repository Librarian, Monash University* Belinda Tiffen - *Director, Library Resources Unit, UTS* Ginny Barbour - *Director, AOASG* Janet Fletcher - *University Librarian, Victoria University, Wellington* Katrina Dewis - *Senior Librarian (Discovery Services) / Acting Associate Director, University of Tasmania* Maude Frances - *Associate Director, Library Digital Repositories, UNSW* Natasha Simons - *Program Lead, Skills Policy and Resources, ANDS* Alexander Sussman - *Associate Director, Academic Services, UNSW Sydney*
CAUL Office Representative	Harry Rolf - *CAUL Communications Officer (Support)*
Program Director	Jill Benn - *University Librarian, University of Western Australia; Program Director for FAIR Access to Research Program*

Document Control:

Filename	**CAUL Project Plan Repositories Submitted Updated Endorsed 20180417 MB.docx**
Author	Jill Benn, Diane Costello, Martin Borchert, Andrew Harrison
Prepared For	CAUL Executive
Status	☑ Draft ☑ Review ☐ Final
Date	17 April 2018

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

PROJECT DESCRIPTION

This project will review the current Australasian open access research output repository infrastructure in order to provide a current assessment and provide recommendations for future development to improve: repository interoperability; compliance with institutional, funder and FAIR policy statements; accessibility and the user experience.

1.1 Background

Australian research repository infrastructure was implemented from 2007 using ASHER funding provided by the Australian Government. This was ground breaking at the time and led to every Australian university benefiting from the availability of an institutional repository, in most cases led by the university library, to facilitate the availability of primarily author final versions of research publications via green open access. Institutions were free to select and implement repository tool(s) of their choice.

Eleven years on, Australian repository infrastructure continues to be diverse in the choice of repository tools used by each institution, and in capabilities, services, and usage. About half of Australian universities have an Open Access (OA) policy or statement, which, together with Australian Research Council (ARC) and National Health and Medical Research Council (NHMRC) - the funders' - OA policies - are major drivers of OA behaviour and compliance today. The policy statement on Findable, Accessible, Interoperable and Reusable (FAIR) Access to Australia's Research Outputs [FAIR policy statement] has developed into a complementary driver for the development and usage of repository infrastructure.

The National Library of Australia's Trove service harvests records of Australian research outputs from institutional research repositories, however there are issues with interoperability, the availability of required metadata for harvesting, comprehensiveness, and the findability and user experience with Australian research outputs being mixed in with a much larger collection of library holdings in Trove.

The project will deliver a number of products addressing each project objective. These products will be in the form of written reports providing a summary of environmental scans, research and recommendations for each objective.

1.2 Objectives

The objectives of this project are to:

1. Review and report on the current Australasian research repository infrastructure.
2. Review and report on the international research repository infrastructure and developments.
3. Develop a series of repository user stories.
4. 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.
6. Investigate and make recommendations for next generation repository tools (consortium approach, possible infrastructure project).

7. Investigate and make recommendations for a possible “Research Australia”

8. Collection of research outputs.

1.3 Benefits

The following benefits have been identified; one for each objective or work package:

- 1. Detail and summarise the current state of the Australasian repository infrastructure to facilitate the gap analysis of the differences between this and the ideal state. The report will also provide a public record of the repository infrastructure as it stood in 2018.

- 2. Assess the current international repository landscape in order to provide relevant advice to CAUL Executive and member institution libraries on necessary steps for national repositories, specifically on international best practice and interoperability with international repositories

- 3.User stories will deliver a comprehensive understanding of current requirements for repository infrastructure across various stakeholder groups, enabling recommendations arising from the project to be user focused and responsive to diverse user groups.

- 4.A roadmap for improving their existing repository infrastructure which will ensure the highest possible level of compliance with FAIR principles using current tools.

- 5. A target for repository interoperability, functionality and usability which can be used for future repository development activities.

- 6. Current information and advice on the preferred strategy for consortium and institutional strategies for selecting and procuring next generation repository tools.

- 7. A ‘Research Australia’ framework outlining a single collection point for all Australian research outputs.

1. SCOPE

The scope of this project is limited to the investigation and provision of reports and recommendations around the application of the FAIR principles to Australasian OA research repository infrastructure, including infrastructure for all research outputs such as research publications, research data and non-traditional research outputs (NTROs). The scope is inclusive of the relationship between Australasian and international infrastructure. The scope may be broadened and extended in future pending a decision point, to progress an RFO, procurement and infrastructure implementation project.

1.1. Out of Scope

- This project will not address Australasian repository infrastructure used for other business purposes such as learning resources including open educational resources (OERs) and digitised library collections.

- This project will not address current research information system (CRIS) or research management systems (RMS), although system integrations are acknowledged.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

- This project will not directly address the use of repositories, CRIS or RMS systems used for the purpose of executing the institutional submission to the Excellence in Research for Australia (ERA) research assessment exercise required by the Australian Research Council (ARC).

- Any implementation of a consortium approach for #6 will be addressed as another follow-on project.

- Any implementation of a "Research Australia" collection for #7 will be addressed as another follow-on project.

2. RISKS AND MITIGATION STRATEGIES

Risk	**Mitigation Strategies**
The project timeline is aspirational and may not be met.	1. Front-loading the timeline may assist. 2. Working on work packages concurrently. 3. An interim report(s) could be provided to CAUL.
Inadequate resources may be available for this project.	1. Project Team appointed from within CAUL Institutions; 2. CAUL Office resources will be prioritised to enable appropriate levels of support; 3. There is the opportunity to incorporate other interest CAUL library employees into the project if required.
The project products (ref. section 7) may not fulfil the requirements of the objectives and the project.	Project products need to be pre-approved by the Program Director prior to work commencing. Project products for objectives #4, #5, #6 and #7 are dependent on the information collected in products for objectives #1, #2, #3. To avoid duplication of effort and unnecessary delays the teams for #1, #2, #3 be scheduled to complete draft versions of their products before June to allow the other teams time to use their products. Project product #6 Next generation repository tools is more work than the other deliverables and potentially the most complex. It needs to be started on immediately upon approval of the work package.
High level of diversity and complexity	CAUL members institutions have varying needs, policies, integrations and priorities for research repository tools. It is likely to be difficult to meet the needs and situation of every institution, resulting in the project products being too general in nature to be of practical use at the institution level. Therefore the project products need to reflect a desired state of enduring principles that the members can work towards over the coming decade as and when they have the resources

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

	and opportunity to upgrade and replace their repository infrastructure.

3. ISSUES

The following issues are identified:

- The repository marketplace, after years of consolidation, is now showing signs of increasing innovation and a new diversity of commercially supported tools. Any work on next generation tools is likely to have a short window of currency.
- Institutions will have diverse needs and ambitions and may not wish to or be able to direct institutional or CAUL resources to completing recommendations.
- The NTRO repository application is not yet well explored and issues around diversity of outputs, format, accessibility etc are likely to become apparent.

4. ESTIMATED RESOURCES

- Personnel to undertake work to meet the seven project objectives (outputs);
- Resources and expertise to develop and deliver use story resources;
- CAUL office personnel re project management (and some of the above);
- Following the delivery of objective / outputs #5 and #6, a decision point will be required of CAUL members and the CAUL Executive to voluntarily opt-in proceed (at institution level) with implementation of next stage project deliverables for a Research Australia collection and next generation repository tools. The cost of implementing the collection and tool(s) may be considerable for the Australian higher education research market.

5. PROJECT MANAGEMENT TEAM STRUCTURE

Program Director:	Jill Benn - *University Librarian at University of Western Australia (UWA); Project Director for FAIR Access to Research program*
Project Lead:	Martin Borchert - University Librarian, UNSW Sydney
Project Team Members:	Andrew Harrison - R*esearch Infrastructure Librarian, Monash University* Belinda Tiffen - *Director, Library Resources Unit, UTS* Ginny Barbour - *Director, AOASG* Janet Fletcher - *University Librarian, Victoria University, Wellington*

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

	Katrina Dewis - *Senior Librarian (Discovery Services) / Acting Associate Director, University of Tasmania* Maude Frances - *Associate Director, Library Repository Services, UNSW Sydney* Natasha Simons - *Program Lead, Skills Policy and Resources, ANDS* Alexander Sussman - *Associate Director (Academic Services), UNSW Sydney*
Project Approval:	CAUL Executive via Program Director
CAUL Office Support:	EO or CPO (which, and how much FTE, to be determined)

6. PROJECT PRODUCTS (Outputs)

(include examples in the template e.g. Environmental Scan, Legal Advice, Communication Plan, Survey Results, Draft Report, Proposal to CAUL, etc.)

Product Title:	**1. Review and report on the current Australasian research repository infrastructure.**
Purpose of product:	Provide an account of the current Australasian research repository infrastructure as a starting point for the project.
Description:	The report will describe current systems available (institutions'), software used, presence of an institutional OA policy, compliance with funder OA policies, extent of the collection, interoperability, and contribution to Trove.
Format & presentation:	PDF MS Word document
Quality:	The document will be reviewed by the project team, project lead, and the Program Director prior to final submission to CAUL Executive.
Authored by:	Lead Andrew Harrison (Monash), Maude Frances (UNSW), Kay Steel (Federation)
Approval required:	Jill Benn, FAIR Program Director and Diane Costello, Executive Officer CAIUL

Product Title:	**2. Review and report on the international research repository infrastructure and developments.**
Purpose of product:	Provide an account of international research repository developments which can be used to benchmark Australasian research repository infrastructure.
Description:	The report will describe the availability and uptake of a range of repository tools, support available, comments on interoperability, funder requirements, and the formation of subject and national repository networks.
Format & presentation:	PDF MS Word document.
Quality:	The document will be reviewed by the project team, project lead, and the Program Director prior to final submission to CAUL Executive.
Authored by:	Lead Ginny Barbour (AOASG), Natasha Simons (ANDS), Scott Abbott (UTS), Alexander Susman (UNSW)
Approval required:	Jill Benn, FAIR Program Director and Diane Costello, Executive Officer CAIUL

Product Title:	**3. Develop a series of repository user stories.**
Purpose of product:	To identify and describe a range of personas and use cases for Australian research repository content and systems, including for research publications (institutional and Research Australia collection), data and NTROs.
Description:	A range of personas and use cases will be identified and described for research publications (institutional and Research Australia collection), data and NTROs. A range of reusable formats will be provided to suit various purposes i.e. author engagement, researcher engagement, procurement, etc.
Format & presentation:	PDF MS Word document, infographics, PPT slides and online perhaps online video interviews.
Quality:	The document will be reviewed by the project team, project lead, and the Program Director prior to final submission to CAUL Executive.

Authored by:	Lead Maude Frances (UNSW), Belinda Tiffen (UTS), Eleanor Colla (UNE)
Approval required:	Jill Benn, FAIR Program Director and Diane Costello, Executive Officer CAIUL

Product Title:	**4. Recommend and make improvements to the current Australasian research repository infrastructure (improve and make the most of what we have).**
Purpose of product:	Provide a list of carefully considered and prioritised improvements that can be made to Australian research repository infrastructure.
Description:	Using the information gained from the "gap analysis" created by products #1, #2 and #3, provide a prioritised list of practical recommendations that can be used to assess and increase repository performance (across the broad range of diverse repository tools) including interoperability, metadata, compliance with funder and institutional OA policies, compliance with the FAIR principles, online services etc.
Format & presentation:	PDF MS Word document.
Quality:	The document will be reviewed by the project team, project lead, and the Program Director prior to final submission to CAUL Executive.
Authored by:	Lead Alexander Sussman (UNSW), Katya Henry (QUT), Margaret Pembroke (SCU)
Approval required:	The document will be reviewed by the project team, project lead, and the Program Director prior to final submission to CAUL Executive.

Product Title:	**5. Develop and report on an ideal state for Australian research repository infrastructure.**
Purpose of product:	To produce a description of an ideal state for research repository infrastructure in Australia, as a target for development, and to facilitate a gap analysis between current and ideal.
Description:	Bringing together knowledge gained from other project products, the purpose of this deliverable is to describe the ideal state for Australian research repositories in terms of a range of factors including

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

	interoperability, metadata, compliance with funder and institutional OA policies, compliance with the FAIR principles, online services etc.
Format & presentation:	PDS MS Word document.
Quality:	The document will be reviewed by the project team, project lead, and the Program Director prior to final submission to CAUL Executive.
Authored by:	Lead Martin Borchert (UNSW) + All Leads
Approval required:	The document will be reviewed by the project team, project lead, and the Program Director prior to final submission to CAUL Executive.

Product Title:	**6. Investigate and make recommendations for next generation repository tools (consortium approach, possible infrastructure project).**
Purpose of product:	Assess the open source and commercial repository tool marketplace to determine options forward for some form of consortial approach to bringing next generation repository(ies) tools to CAUL consortium members.
Description:	Use the environmental scan in project product #2 to develop a greater understanding of the features, costs and benefits of various open source and commercially supported (may be open source or proprietary) repository tools. Develop an agreed set of repository tool requirements and test the marketplace via an RFI process. Take this information to CAUL Executive (and CAUL) to enquire whether there is any appetite for a consortial approach to procurement, implementation and support for research repository tools at the institutional level. The information gained may also be used to inform project product #5 to recommend a solution for hosting and providing FAIR access to the Research Australia collection.
Format & presentation:	PDF MS Word document.
Quality:	The document will be reviewed by the project team, project lead, and the Program Director prior to final submission to CAUL Executive. NLA, COAR, also suggest SHARE, OpenAIRE and La Referencia be consulted.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Authored by:	Lead Martin Borchert UNSW), Janet Fletcher (VUW), Elke Dawson (ANU)
Approval required:	The products will be reviewed by the project team, project lead, and the Program Director prior to final submission to CAUL Executive.

Product Title:	**7. Investigate and make recommendations for a possible "Research Australia" collection of research outputs.**
Purpose of product:	Investigate marketplace demand for the collection of Australian research outputs, assess against the current Trove collection and identify a range of options meeting the FAIR principles.
Description:	Provide an assessment of the effectiveness of the Trove and Research Data Australia collections of Australian research repository outputs and assess a gap analysis with respect to the ideal state. Consider options working with the NLA to improve Trove to provide a "Research Australia" collection, and also consider other options for creating such a collection. Make recommendations on the preferred path forward which will provide global access to the nation's collective research outputs, while meeting the FAIR principles and expectations of digital client experience and usability.
Format & presentation:	PDF MS Word document.
Quality:	The document will be reviewed by the project team, project lead, and the Program Director prior to final submission to CAUL Executive. NLA, ANDS, COAR, also suggest SHARE, OpenAIRE and La Referencia be consulted.
Authored by:	Lead Katrina Dewis (UTAS), Belinda Tiffen (UTS), NLA representative (suggest Julia Hickie)
Approval required:	The document will be reviewed by the project team, project lead, and the Program Director prior to final submission to CAUL Executive.

7. PROJECT TIMELINE

8. Start Date: 15 March 2018

9. Expected Completion Date: 14 September 2018

10. Detailed project product timeline

Dates indicate end of month.

Objective (work package)	**Start Date**	**Completion Date, submission to Program Director**	**CAUL Approval / Decision Point Date**
1. Review and report on the current Australasian research repository infrastructure.	April 2018	May 2018	June
2.Review and report on the international research repository infrastructure and developments.	April	May	June
3.Develop a series of repository user stories.	April	May	July
4. Recommend and make improvements to the current Australasian research repository infrastructure (improve and make the most of what we have)	April	July	August
5.Develop and report on an ideal state for Australian research repository infrastructure	April	July	September
6.Investigate and make recommendations for next generation repository tools (consortium approach, possible infrastructure project).	April	August	September
7.Investigate and make recommendations for a possible “Research Australia” collection of research outputs.	April	August	September

11. STAKEHOLDER ENGAGEMENT AND COMMUNICATION PLAN

Stakeholders:

- Project team members

- Program Director, Fair, affordable and open access to knowledge program
- CAUL Council members
- CAUL Executive
- CAIRSS repository community
- ARC and NHMRC
- Repository tool vendors and communities

11.1. Reporting Requirements (to whom and frequency)

11.1.1. Reporting to Program Director

Brief monthly progress reports will be provided to Jill Benn, Program Director.

See Section 12 Detailed project product timeline for delivery dates of reports.

11.2. Reporting to CAUL/Public (Newsletter)

See Section 12 Detailed project product timeline for delivery dates of reports.

Two news items will be delivered via the CAUL Newsletter, in July and October.

Project products will be made available on open access via the CAUL website and an OA repository.

12. RELATED PROJECTS OR INITIATIVES

The CAUL review of Australian repository infrastructure project is one of five projects under the CAUL Fair, affordable and open access to knowledge program.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Appendix 3: Summary table of Infrastructure/Tools/Supporting Organisations

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

Name (Location)	Infrastructure/ Supporting Organisations/ Tool	Short description
La Referencia (Latin America)	Infrastructure	La Referencia, is a Latin American network of open access repositories. Through its services, it supports national Open Access strategies in Latin America through a platform with interoperability standards, sharing and giving visibility to the scientific production generated in institutions of higher education and scientific research.
OpenAIRE (Open Access Infrastructure for Research in Europe)	Infrastructure	OpenAIRE is a massive and successful infrastructure project that is well funded by an organised EU. Lessons can be learnt from the centralised top down funding and management of such a complex and successful project. RIOXX to be folded into the Open AIRE metadata schema.
Open DOAR (International)	Infrastructure	*Open*DOAR is an authoritative directory of academic open access repositories.
PubMedCentral - PMC (US)	Infrastructure	PMC is an American free archive/repository for biomedical and life sciences journal literature deposited by participating journals, as well as for author manuscripts that have been submitted in compliance with the public access policies of participating research funding agencies.
SHARE - SHared Access Research Ecosystem (US)	Infrastructure	As an established platform with back-end and front-end source code (for the research database and the platform) open source, SHARE is a structure worth looking at more closely. SHARE adapts to the data sources' metadata structure and to refine this process they have recently published the SHARE Version 3 Metadata Harvesting update. Current apparent lack of

		connection between Open AIRE and SHARE, however, needs to be considered.
CORE (UK)	Tool	CORE (Connecting Repositories) is a UK based free service aimed at aggregating all open access content distributed across different systems.
Dataverse (International)	Tool	Dataverse is an open source web application to share, preserve, cite, explore, and analyse research data. A Dataverse repository is a software installation, which then hosts multiple virtual archives called Dataverses. Each dataverse contains datasets, and each dataset contains descriptive metadata and data files (including documentation and code that accompany the data).
Dryad (International)	Tool	Dryad is an international disciplinary repository for data underlying scientific and medical publications. Dryad is open source DSpace repository software and advocates for making data Findable, Accessible, Interoperable, and Reusable.
Figshare (International)	Tool	Figshare is a privately owned company. Cutting edge proprietary system to surface open research content (data and publications) from a wide variety of sources. Very attractive in terms of a ready made solution. Their publisher model requires an SLA statement guaranteeing (only) 10 years of persistent availability. (From here)
IRUS-UK (Institutional Repository Usage Statistics United Kingdom)	Tool	A repository plugin which provides COUNTER-conformant usage statistics for all content downloaded from participating UK institutional repositories (IRs). There is currently an Australian trial underway.
Re3data (Registry of Research Data Repositories)	Tool	re3data.org is a global registry of research data repositories from different academic disciplines. It is worth noting that re3data will be used in the AGU FAIR project as the basis of a tool researchers can use that guides them to select which repository to best deposit their data in.
ResourceSync	Tool	ResourceSync is an ANSI/NISO specification that describes a synchronization framework for the web consisting of various capabilities that allow third-party systems to remain synchronized with a server's evolving resources.

RIOXX (Metadata Application Profile United Kingdom)	Tool	RIOXX is a metadata application profile that provide a mechanism to help UK institutional repositories comply with the RCUK policy on open access
Scholix initiative (Scholarly Link Exchange)	Tool	Scholix is a work in progress from Research Data Alliance; model to link publications with data; will have global application
CARL (Canadian Association of Research Libraries)	Supporting Organisation	CARL has produced a repositories roadmap and national Open Access and Open Science infrastructure supporting the research agenda.
COAR (Confederation of Open Access Repositories) + Next Generation Repositories Project	Supporting Organisation	COAR is an international association with over 100 members and partners from around the world representing libraries, universities, research institutions, government funders and others. COAR brings together the repository community and major repository networks in order build capacity, align policies and practices, and act as a global voice for the repository community.
LERU (Europe)	Supporting Organisation	The League of European Research Universities (LERU) is a well-established network of research-intensive universities.
LIBER -Ligue des Bibliothèques Européennes de Recherche (Europe)	Supporting Organisation	LIBER is a European network of more than 400 national, university and other libraries.
National Institute of Informatics (NII), Japan	Supporting Organisation	The NII is an inter-university Japanese research institute for advancing the study of informatics. Research at NII focuses on information-gathering techniques and systems for information management. The NII in its push for open access and open science drives the development of academic information infrastructures.
SPARC	Supporting Organisation	SPARC (the Scholarly Publishing and Academic Resources Coalition) is a US based advocacy group.
UK-CORR (United Kingdom Council of Research Repositories)	Supporting Organisation	The professional organisation for UK open access repository administrators and managers.

Details of Infrastructure/Tools/Supporting Organisations

Infrastructure

Name	La Referencia
Description	SHARING AND GIVING VISIBILITY TO SCIENTIFIC PRODUCTION IN LATIN AMERICA The Federated Network of Institutional Repositories of Scientific Publications, or simply LA Referencia, is a Latin American network of open access repositories. Through its services, it supports national Open Access strategies in Latin America through a platform with interoperability standards, sharing and giving visibility to the scientific production generated in institutions of higher education and scientific research. Note for participation they say “BE PART OF LA REFERENCIA LA Referencia does not harvest directly from the repositories but through the national nodes. Institutions / repositories that want to be part of the network should contact the representative of their countries and follow the technical recommendations of integration to the national node. If your country wishes to participate as a national node, please contact the Executive Secretariat.”
Funding	“We Have A Support Network Sparked By All Latin America” Presumably by funding from member organisation but not explicit.
Comments	Partnerships with Open Aire and COAR Very visible internationally.

Name	OpenAIRE - Open Access Infrastructure for Research in Europe https://www.openaire.eu/
Description	OpenAIRE is a network of Open Access repositories, archives and journals that support Open Access policies. It goes beyond the traditional publications aggregator by interconnecting entities related to scholarly communication (publications, research data, funding, people, organizations, data sources) allowing users to navigate alongside a rich information space graph and provides a wide range of services, from deposition to statistics. OpenAIRE has started out as a policy support mechanism for the EC (FP7 pilot and H2020 OA policies), with the aim to be *the* European

	scholarly communication hub providing its services to many European funders. https://www.openaire.eu/supporting-open-science "50 partners, from all EU countries, and beyond, will collaborate to work on this large-scale initiative that aims to promote open scholarship and substantially improve the discoverability and reusability of research publications and data. The initiative brings together professionals from research libraries, open scholarship organisations, national e-Infrastructure and data experts, IT and legal researchers, showcasing the truly collaborative nature of this pan-European endeavor. A network of people, represented by the National Open Access Desks (NOADs), will organise activities to collect H2020 project outputs, and support research data management. Backing this vast outreach, is the OpenAIRE platform, the technical infrastructure that is vital for pulling together and interconnecting the large-scale collections of research outputs across Europe. The project will create workflows and services on top of this valuable repository content, which will enable an interoperable network of repositories (via the adoption of common guidelines), and easy upload into an all-purpose repository (via Zenodo). OpenAIRE2020 will assist in monitoring H2020 research outputs and will be a key infrastructure for reporting H2020's scientific publications as it will be loosely coupled to the EC's IT backend systems. The EC's Research Data Pilot will be supported through European-wide outreach for best research data management practices and Zenodo, which will provide long-tail data storage. Other activities include: collaboration with national funders to reinforce the infrastructure's research analytic services; an APC Gold OA pilot for FP7 publications with collaboration fromLIBER; novel methods of review and scientific publishing with the involvement of hypotheses.org; a study and a pilot on scientific indicators related to open access with CWTS's assistance; legal studies to investigate data privacy issues relevant to the Open Data Pilot; international alignment with related networks elsewhere with the involvement of COAR." (From here)
Software (note open or closed)	Information not easily available: CRIS - "loosely coupled" with the EC's IT backend systems?
Metadata requirements	OpenAIRE expects metadata to be encoded in the Dublin Core metadata format (metadataPrefixoai_dc). For information on how to use the individual DC fields, please refer to the section "Use of OAI-DC" below. (From here) See Crosswalk from RIOXX to OpenAire here. Note: there is speculation as to whether RIOXX might be incorporated into new OpenAIRE guidelines and schema but this is unclear.
Recommendations for interoperability	There are OpenAIRE guidelines for interoperability between OpenAIRE infrastructure (itself a CRIS system) and 1. literature repositories (OAI-PMH and OAI-DC), 2. data archives (OAI-PMH and

	Datacite) and 3. other CRIS systems. For more information search OpenAIRE Guidelines here.
Funder requirements	Under Horizon 2020, each beneficiary must ensure open access to all peer-reviewed scientific publications relating to its results. (See here)
Policy requirements	https://www.openaire.eu/oa-policies-mandates
Key relationships	The EU member states, La Referencia, many non-European publication, data and CRIS sources
Funding	OpenAIRE is an EC funded project (FP7 246686 and 283595)
Compliance with FAIR principles	In part. The EU Horizon 2020 policy expects: "Under Horizon 2020, each beneficiary must ensure open access to all peer-reviewed scientific publications relating to its results. To meet this requirement, beneficiaries must, at the very least, ensure that any scientific peer-reviewed publications can be read online, downloaded and printed. Since any further rights - such as the right to copy, distribute, search, link, crawl and mine - make publications more useful, beneficiaries should make every effort to provide as many of these options as possible." (From here)
Support available/	OpenAIRE Helpdesk
Notes	OpenAIRE is a massive and successful infrastructure project that is well funded by an organised EU. Lessons can be learnt from the centralised top down funding and management of such a complex and successful project. Research gaps: could not find information on the nature of the OpenAIRE infrastructure or organisation itself. RIOXX to be folded into the Open AIRE metadata schema.

Name	Open DOAR http://www.opendoar.org/ - moving to new site, currently in alpha http://v2.opendoar.sherpa.ac.uk/
Description	*Open*DOAR is an authoritative directory of academic open access repositories. Each *Open*DOAR repository has been visited by project staff to check the information that is recorded here. This in-depth approach does not rely on automated analysis and gives a quality-controlled list of repositories.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

	*Open*DOAR has also been identified as a key resource for the Open Access community (K.B.Oliver & R.Swain, 2006 - PDF) and identified as the leader in repository directories in a study by Johns Hopkins University. *Open*DOAR was one of the services which contributed to SHERPA being awarded the 2007 SPARC Europe Award for Outstanding Achievements in Scholarly Communications.
Funding	*Open*DOAR is one of the SHERPA Services including RoMEO and JULIET, run by the Centre for Research Communcations (CRC). Current development work is currently funded by JISC, with contributions from the CRC host organisation, the University of Nottingham.
Comments	Has a policy tool too - which includes use of metadata (though does not define what the metadata should be) http://v2.opendoar.sherpa.ac.uk/policytool/ Doesn't have way I can see of highlighting networks of repos.

Name	PubMed Central (PMC) https://www.ncbi.nlm.nih.gov/pmc/
Description	PMC is a free archive/repository for biomedical and life sciences journal literature deposited by participating journals, as well as for author manuscripts that have been submitted in compliance with the public access policies of participating research funding agencies. https://www.ncbi.nlm.nih.gov/pmc/about/intro/
Software (note open or closed)	PMC is not software in the strict sense of the word, however the archive is free and interoperable.

Metadata requirements	A journal must provide PMC with the **full text of articles** in an XML format that conforms to an acceptable journal article DTD (Document Type Definition) and meets the PMC Minimum Data Criteria. PMC does not accept articles in HTML format. NLM recommends that data be submitted in XML conforming to the NISO JATS Journal Publishing Tag Set, but PMC will also accept data in other full-text article DTDs that are widely used in life sciences journal publishing. Information for publishers is available from https://www.ncbi.nlm.nih.gov/pmc/pub/pubinfo/
Recommendations for interoperability	

Funder requirements	PMC serves as the full-text repository for papers across a variety of scientific disciplines that fall under several funding agency public access policies. PMC archived papers need to comply with the following agencies: ● National Institutes of Health (NIH) ● Administration for Community Living (ACL) [PDF] ● Agency for Healthcare Research and Quality (AHRQ) ● Bill & Melinda Gates Foundation ● Centers for Disease Control and Prevention (CDC) ● Environmental Protection Agency (EPA) ● Food and Drug Administration (FDA) ● Howard Hughes Medical Institute (HHMI) ● National Aeronautics and Space Administration (NASA) ● National Institute of Standards and Technology (NIST) ● Office of the Assistant Secretary for Preparedness and Response (ASPR) ● U.S. Department of Homeland Security ● U.S. Department of Veterans Affairs (VA) https://www.ncbi.nlm.nih.gov/pmc/about/public-access/

Policy requirements	Policies related to participation, article types and content, and article availability and usage can be found at: https://www.ncbi.nlm.nih.gov/pmc/about/guidelines/
Key relationships	U.S. National Institutes of Health's national Library of Medicine; U.S. National Library of Medicine's National Centre for Biotechnology Information (NCMI).
Compliance with FAIR principles	Policies relate to systematic downloading of articles, not FAIR use of PMC records and metadata
Support available/	Support is initially via email: nihms-help@ncbi.nlm.nih.gov Support is then directed towards authors, publishers, users etc.
Notes	

Name	SHARE http://www.share-research.org/
Description	"SHARE is a higher education initiative whose mission is to maximize research impact by making research widely accessible, discoverable, and reusable. To fulfil this mission SHARE is developing services to gather and freely share information about research and scholarly activities across their life cycle. Making research and scholarship freely and openly available encourages innovation and increases the diversity of innovators. Where open metadata about research already exists, its usefulness is limited by poor or inconsistent quality or by difficulty of access. For most individuals or groups to use this data, the cost of accessing, collecting, and improving the data is too great. SHARE is building its free, open, data set by gathering, cleaning, linking, and enhancing metadata that describes research activities and outputs—from data management plans and grant proposals to preprints, presentations, journal articles, and research data." (From here)
Software (note open or closed)	Backend source code Open on Github here: https://github.com/CenterForOpenScience/SHARE Frontend source code Open on github here:

	https://github.com/CenterForOpenScience/ember-share
Metadata requirements	Apparently SHARE adapts to harvest whatever metadata schema a source provides. See here: https://osf.io/wur56/wiki/Metadata%20Analysis/ and the SHARE FAQ's on metadata for other links: https://osf.io/cpsin/wiki/home/ and SHARE Version 3 Metadata Harvesting update
Recommendations for interoperability	"SHARE values **international interoperability**. As a North American initiative, we realize that we represent only a portion of the research effort across the globe. We look for opportunities to develop and adopt standards and practices that help knit the world's research together into a more cohesive and accessible whole." (From here)
Funder requirements	SHARE is not a funder but a funded system-architecture to harvest repositories and publishers so as to provide access to open research content in one place. "SHARE was originally funded by the Institute of Museum and Library Services (IMLS) and the Alfred P. Sloan Foundation, and the core technology was built by the Center for Open Science (COS) in collaboration with the Association of Research Libraries (ARL). The SHARE initiative was founded in 2013 by ARL, the Association of American Universities (AAU), and the Association of Public and Land-grant Universities (APLU)." (From here)
Policy requirements	While not a policy page this SHARE FAQ page provides some useful information and links as a starting point.
Key relationships	http://www.share-research.org/projects/projects-and-partners/
Compliance with FAIR principles	SHARE harvests metadata from 169 minor, middle and major sources from arxiv (preprint server) to many institutional repositories including the University of Wollongong and many US-based IRs. It also harvests major publishers such as Springer and others. As a harvester and portal it provides the user the ability to locate papers and other research outputs so the works are Findable and Accessible and, depending on the licenses on those individual outputs at their original location, the works will be more or less Interoperable and Reusable.
Support available/ Could be person	About Jeff Spies at SHARE Jeff Spies website CV and contact details Or others at http://www.share-research.org/about/our-team/
Notes	As an open source system with some great work already done and being done would it be possible to use SHARE's open code as a starting point for an Australian open research portal that replicates much of SHARE's dataset

	of sources while specifically focusing on connecting and surfacing the content of Australian IRs to the world? Australasian funding could significantly boost the work still being done by the US-SHARE team and could help bridge the Australasian and North American open communities. Linking these with OpenAire, LaReferencia and others would be a huge win for the open movement in my opinion and would avoid the replication of work we don't need to see more of. The existing flexible metadata harvesting system seems very promising and as mentioned above and here, more work is progressing in this area. Open source code is critical and available for both the SHARE back and front end. Links in the "Software" field of this table.

Tools

Name	CORE https://core.ac.uk/
Description	CORE (Connecting Repositories) is a free service aimed at aggregating all open access content distributed across different systems. CORE does this by harvesting data providers including institutional repositories, subject-repositories and journal publishers. CORE provides text and data mining access through supporting programmable machine access to the content, for example use of API or data dumps.
Software (note open or closed)	https://core.ac.uk/images/bb77d4f.pdf
Metadata requirements	We mainly support oai_dc, the mainstream metadata format used in the OAI-PMH Protocol, utilising the Dublin Core vocabulary, a popular vocabulary for bibliographic data. We also support RIOXX, a richer metadata protocol, used mostly by the UK repositories.
Recommendations for interoperability	CORE API

Funder requirements	
Policy requirements	
Key relationships	Open University Jisc There are partner projects with OpenAIRE, COAR, IRUS-UK, ORCiD, RIOXX
Compliance with FAIR principles	Yes
Support available/ Could be person	theteam@core.ac.uk
Notes	To discuss project and research collaborations or if you represent an industry partner interested in using CORE in your work, please contact: Dr. Petr Knoth Knowledge Media institute, The Open University Walton Hall, Milton Keynes, MK7 6AA, BUCKS, United Kingdom petr.knoth@open.ac.uk

Name	Dataverse Full name: The Dataverse Project https://dataverse.org/

Description	Dataverse is an open source web application to share, preserve, cite, explore, and analyze research data. A Dataverse repository is the software installation, which then hosts multiple virtual archives called Dataverses. Each dataverse contains datasets, and each dataset contains descriptive metadata and data files (including documentation and code that accompany the data). As an organizing method, dataverses may also contain other dataverses. https://dataverse.org/about
Software (note open or closed)	Fully open source and as such there are several Dataverse repositories installed in universities and organisations around the world. E.g. ADA the Australian Data Archive https://dataverse.ada.edu.au/
Metadata requirements	Dataverse uses standard-compliant metadata to ensure that Dataverse metadata can be mapped easily to standard metadata schemas and be exported into JSON format (XML for tabular file metadata) for preservation and interoperability. Dataverse metadata is compliant with DDI Lite, DDI 2.5, Codebook, DataCite 3.1, and Dublin Core amongst others. http://guides.dataverse.org/en/latest/user/appendix.html
Recommendations for interoperability	The Dataverse currently has multiple open APIs available, which allow for searching, depositing and accessing data. Dataverse is integrated into Open Science Framework and Open Journal Systems, amongst others. The dataverse developers information provides links to the development community. The Dataverse Development Community is an active group of internal and external contributors to the Dataverse software codebase. https://dataverse.org/developers
Funder requirements	Funded by Harvard with additional support from the Alfred P. Sloan Foundation, National Science Foundation, National Institutes of Health, Helmsley Charitable Trust, IQSS's Henry A. Murray Research Archive, and many others.
Policy requirements	Dataverse policies are available from https://dataverse.org/best-practices/harvard-dataverse-policies

Key relationships	Dataverse is also installed in the EU in the Netherlands, Germany, France and Finland. The largest Dataverse repository is DataverseNL in the Netherlands providing data management services for 11 Dutch Universities. A similar service is being developed in Norway.
Compliance with FAIR principles	Not a great deal of information however the Dataverse Project does have an available presentation on FAIR Data Management and FAIR Data Sharing at https://dataverse.org/presentations/fair-data-management-and-fair-data-sharing
Support available/ Could be person	https://dataverse.org/contact

Name	Dryad https://www.datadryad.org/
Description	Dryad is an international disciplinary repository for data underlying scientific and medical publications. Dryad facilitates the data archiving, preferably at the time of publication when the data are readily available. Dryad curators review submitted data and perform quality control on metadata before inclusion of new content in the repository. Approved data is allocated to specialized repositories to exchange identifiers and other metadata for cross-referencing.
Software (note open or closed)	Open source DSpace repository software.
Metadata requirements	Dublin Core
Recommendations for interoperability	Not applicable

Funder requirements	Not-for-profit membership
Policy requirements	Policy requirement for accepting data are located at https://www.datadryad.org/pages/policies By default, all accepted content is made public upon publication.
Key relationships	Metadata Research Centre http://cci.drexel.edu/mrc/ DMCI Science and Metadata Community http://dublincore.org/groups/sam/ DataCite for Digital Object Identifiers. California Digital Library - new relationship announced in 2018.
Compliance with FAIR principles	Dryad advocates for making data Findable, Accessible, Interoperable, and Reusable. All files submitted to Dryad must abide by the terms of Creative Commons Zero (CC0 1.0). Terms of use require data to be in non-proprietary open formats.
Support available/ Could be person	Dryad PO Box 585 Durham, NC 27702-0585 USA

Name	Figshare https://figshare.com/
Description	Initially a start-up, figshare is now part of https://www.digital-science.com/ which is owned by **Holtzbrinck Publishing Group** - a privately-held Stuttgart-based company which owns publishing companies worldwide. “Figshare is a repository where users can make all of their research outputs available in a citable, shareable and discoverable manner”. (From here) Please also read this relevant Figshare blogpost about Figshare’s move into the “next gen” IR domain: “Announcing the figshare Institutional Repository...and Data Repository...and Thesis Repository...really just an all-in-one next gen repository”

Software (note open or closed)	Closed
Metadata requirements	**"Do you [Figshare] support custom metadata?** Yes, you are able to add custom metadata fields at an institutional or department level. (From here) https://figshare.com/services/institutions
Recommendations for interoperability	**"Can figshare integrate with our institutional repository, CRIS and RIMS?** All research on figshare can be pushed to any institutional repository eg. Dspace, Eprints Fedora. We can also integrate with any RIMS or CRIS eg. Symplectic or Pure." (From here) https://figshare.com/services/institutions
Funder requirements	Figshare is a research repository system for individuals, publishers and institutions. Could not find any information on funder requirements beyond this Q & A: **"[Is Figshare] compliant with funder and legal requirements?** We allow a gated publication process to ensure any public facing research is compliant with institutional, funder and legal requirements" (From here) https://figshare.com/services/institutions
Policy requirements	
Key relationships	Digital Science stable of products at https://www.digital-science.com/ Monash University, University of Melbourne, other universities, large publishers, individual users, etc.
Compliance with FAIR principles	"Figshare is building out our existing data (or non-traditional research outputs) repository functionality to cover all the functionality needed for a traditional Institutional Repository (IR). We have worked hard over the past 5 years to build what we believe to be the best data repository available. We have worked on distributing our infrastructure to comply with funder storage policies around the globe, whilst consistently adding functionality, from citation counts to customizable metadata. As such, we now have a customizable system that can live on institutional domains, with institutional branding and institutional DOIs. We accept any file format and are focussed on making all outputs findable, accessible, interoperable and reusable (FAIR) for humans and machines." (From here)
Support available/ Could be person	info@figshare.com

Notes	Privately owned company. Cutting edge proprietary system to surface open research content from a variety of sources. Very attractive in terms of a readymade solution. **"How long will figshare host and retain my public research data for?** Items will be retained for the lifetime of the repository. figshare has been working hard to establish a business model that supports sustainability of the research outputs hosted on figshare. Our publisher model requires an SLA statement guaranteeing 10 years of persistent availability." (From here) https://knowledge.figshare.com/articles/item/how-persistent-is-my-research

Name	Institutional Repository Usage Statistics UK (IRUS-UK)
Description	IRUS-UK is a Jisc-funded UK national aggregation service, which provides COUNTER-conformant usage statistics for all content downloaded from participating UK institutional repositories (IRs). http://irus.mimas.ac.uk/ IRUS-UK enables partnering universities to share and expose statistics of UK repository usage based on the COUNTER standard.
Software (note open or closed)	Data is gathered using a plugin developed for EPrints, DSpace and Fedora. Output of COUNTER statistics is via reports in JUSP in HTML, CSV or TSV formats. Software platforms in use by partnering universities include DSpace, Pure, EPrints, Fedora, Worktribe, and "other" (http://irus.mimas.ac.uk/about/participants/).
Metadata requirements	Not applicable
Recommendations for interoperability	Currently plugins have been developed only with EPrints, DSpace and Fedora repositories. This accounts for two thirds of all UK repositories.

Funder requirements	Funded by Jisc to July 2018 (Jisc is a not-for-profit provider of digital services and solutions including operating shared digital infrastructure and negotiating consortia deals with IT vendors and commercial publishers). 2018/2019 Jiscl budget has not been confirmed. http://irus.mimas.ac.uk/support/faqs/
Policy requirements	Not applicable
Key relationships	The services are being developed by a consortium involving Jisc, Cranfield University and Evidence Base.
Compliance with FAIR principles	
Support available	Details not provided on the website help@jisc.ac.uk Jisc has offices in 7 UK locations London Brettenham House 5 Lancaster Place London WC2E 7EN
Notes	We found the plugin really useful when he was working in the UK because of the ability for consistent statistics to be used for benchmarking usage. Robin gave a presentation on this at Open Repositories 2017 and lobbied for Australian participation. The University of Sydney has since installed the plugin. IRUS-AUS is supported by CAVAL.

Name	**r3data** https://www.re3data.org/
Description	re3data.org is a global registry of research data repositories that covers research data repositories from different academic disciplines. It presents repositories for the permanent storage and access of data sets to researchers, funding bodies, publishers and scholarly institutions. re3data.org promotes a culture of sharing, increased access and better visibility of research data. The registry went live in autumn 2012 and is funded by the German Research Foundation (DFG). (From here)
Software (note open or closed)	Open: Yes it is possible to re-use the re3data metadata. You can fetch the datasets via API (From here)
Metadata requirements	Research data repository (RDR) operators can suggest their infrastructures to be listed in re3data via an application form providing the name and URL as well as other properties of the RDR . The project team thoroughly analyses the website of the RDR using a handbook that gives practical information on how to obtain the metadata properties of the re3data schema. A repository is indexed when the minimum requirements of the re3data policy are met, meaning that mode of access to the data and repository as well as the terms of use must be clearly explained on the repository web sites and the repository must have a focus research data. Before a new record of a RDR is published in re3data all gathered information is reviewed by a second team member (see figure below). Afterwards the RDR will be visible to the public. (From here)
Recommendations for interoperability	See https://www.re3data.org/faq
Funder requirements	
Policy requirements	The use of re3data.org is also recommended in the European Commission's "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020". (From here)

Key relationships	Project partners in re3data.org are the Berlin School of Library and Information Science at the Humboldt-Universität zu Berlin, the Library and Information Services department (LIS) of the GFZ German Research Centre for Geosciences, the KIT Library at the Karlsruhe Institute of Technology (KIT) and the Libraries of the Purdue University. The German partners are actively involved in the German Initiative for Network Information (DINI) and current research data management activities. (From here)
Compliance with FAIR principles	Aiming to help make research data in various research data repositories (RDR's) FAIR. (see under 'notes' re: AGU FAIR project)
Support available/ Could be person	Chairmen: Frank Scholze, Director of KIT Library Michael Witt, Head of Distributed Data Curation Center at Purdue University Libraries email: info@re3data.org
Notes	re3data.org is a global registry of research data repositories that covers research data repositories from different academic disciplines. It is worth noting that re3data will be used in the AGU FAIR project as the basis of a tool researchers can use that guides them to select which repository to best deposit their data in

Name	ResourceSync
Purpose	ResourceSync is an ANSI/NISO specification that describes a synchronization framework for the web consisting of various capabilities that allow third-party systems to remain synchronized with a server's evolving resources. All capabilities are implemented on the basis of the document formats introduced by the Sitemap protocol.
Advantages	Repositories of scholarly articles and data have typically shared metadata via OAI-PMH. As these repositories are re-architected to become resource- or web-centric, the ResourceSync Framework enables sharing of both metadata and content with aggregators and commodity web search engines alike. ResourceSync provides a standard synchronization method that will reduce implementation effort and facilitate easier reuse of resources.

Drawbacks	OAI-PMH is still the main standard for exchanging m-2-m metadata between repositories globally.
Uptake / availability	Unclear, early days.
Benefits for Australian IRs	As for all repositories globally (see advantages).
Steps to participate	Unclear, early days.
Notes	Suggest that a group of technical oriented repository managers from Australia/New Zealand follow this up via direct contact with the Resource Sync folks and also with the National Library of Australia and the ARDC since they are the ones who currently harvest publications and datasets (respectively) from our institutional repositories.

Name	RIOXX
Purpose	This is a metadata application profile that provide a mechanism to help UK institutional repositories comply with the RCUK policy on open access.
Advantages	Applies consistency of metadata used to record research funder and project/grant identifiers and make consistent tracking of open-access research output.
Drawbacks	The future of RIOXX itself is uncertain as it has not been funded for many years. Paul Walk, who maintains RIOXX in his spare time, is likely to be contacted by OpenAIRE to help develop their guidelines and application profile. If that goes ahead, those guidelines are likely to replace RIOXX. Overall, the future of RIOXX is unclear.
Uptake / availability	50% of UK.

Benefits for Australian IRs	Adoption of consistent metadata definitions/practices already developed.
Steps to participate	Unclear [see notes in drawbacks].

Name	Scholix
Purpose	The RDA/WDS Scholarly Link Exchange (Scholix) WG is a follow-up group from the RDA/WDS Publishing Data Services WG. The group was endorsed by the both the RDA and the scientific committee of ICSU-WDS, and is going to be completed later in 2018. The Scholix WG's aim was to enable a comprehensive global view of the links between scholarly literature and data. The working group has leveraged existing work and international initiatives to work towards a global information commons by establishing: ● Pathfinder services and enabling infrastructure ● An interoperability framework with guidelines and standards (see also www.scholix.org) ● A significant consensus ● Support for communities of practice and implementation One important aspect of the Scholix WG is that it has been driven by multiple organisations from different communities from the start, including PID service providers (DataCite, CrossRef, ANDS), publishers (Elsevier, Springer Nature, Clarivate), and repositories (Pangea, European Bioinformatics Institute (EMBL-EBI). ANDS has been a key contributor and sponsor of the Scholix initiative.
Advantages	Uses existing infrastructure, no additional configuration or tools required beyond contributing standard metadata.
Drawbacks	Where the links are not already captured in the repository, it will be additional work to include the publication-data links in metadata provided to a Scholix hub.
Uptake/availability	Major publishers and repositories are committed to this initiative (see purpose). The Scholix WG has made great inroads on its aims. Currently, Scholix maintains an evolving set of Guidelines consisting of:

	(i) an information model (conceptual definition of what is a Scholix scholarly link), (ii) a link metadata schema (set of metadata fields representing a Scholix link), and (iii) a corresponding XML and JSON schema.
Benefits for Aus IRs	Increase exposure of repository content (data and publications) at a global level - in publisher discovery services such as Scopus and in other Scholix Hubs including DataCite, CrossRef and OpenAire.
Steps to participate	Enhance your metadata feed to one of the Scholix 'hubs'. In Australia, simply provide these links in metadata records harvested by RDA and/or provide them to DataCite when minting a DOI through the ANDS service. See guide 'Scholix for institutional repository managers' and info on ANDS website.

Supporting organisations

Name	CARL (Canadian Association of Research Libraries).
Description	
Funding	Funded by institutional library membership.
Comments	Members include Canada's twenty-nine largest university libraries and two federal institutions.CARL provides leadership on behalf of Canada's research libraries and enhances capacity to advance research and higher education. It promotes effective and sustainable knowledge creation, dissemination, and preservation, and public policy that enable broad access to scholarly information. Notable initiative is The Portage network - dedicated to the shared stewardship of research data in Canada.

Name	Confederation of Open Access Repositories (COAR).
Description	COAR is an international association that brings together the repository community and major repository networks in order build capacity, align policies and practices, and act as a global voice for the repository community.
Software (note open or closed)	Not specified however COAR advocates for open access repositories.
Metadata requirements	Working group formed in 2014 to develop a blueprint to outline steps needed to ensure greater interoperability across repository networks. COAR controlled vocabularies for Repository Assets - https://www.coar-repositories.org/files/coar_resourcetypes_v1_0.pdf
Recommendations for interoperability	https://www.coar-repositories.org/files/Roadmap_final_formatted_20150203.pdf
Funder requirements	Association membership
Policy requirements	Articles of association: https://www.coar-repositories.org/about/coar-ev/articles-of-association/
Key relationships	Consortia Advancing Standards in Research Administration Information (CASRAI). La Referencia
Compliance with FAIR principles	Not specified on the website

Support available	Katharina Müller c/o Göttingen State and University Library Platz der Göttinger Sieben 1 37073 Göttingen Germany Phone: +49 551 39-22215 Fax: +49 551 39-5222 E-Mail: office@coar-repositories.org Kathleen Shearer, Executive Director
Notes	

Name	LERU (Ligue des Bibliothèques Européennes de Recherche)
Description	The League of European Research Universities (LERU) is a well-established network of research-intensive universities.
Funding	Funded by institutional library membership.
Comments	They say “We aim at furthering politicians’, policy makers’ and opinion leaders’ understanding of the important role and activities of research-intensive universities. Our 23 members bring together representatives to work on LERU policy development and engage in mutual learning in many areas.” Repositories are only part of a much broader remit for this group. Important advocacy group with Rectors’ involvement as well as other senior university executives.

Name	LIBER (Ligue des Bibliothèques Européennes de Recherche)
Description	LIBER - “the voice of Europe’s library research community” More than 400 national, university and other libraries are part of LIBER and their wider network includes goal-oriented partnerships with other organisations in Europe and beyond. Their 2018-2022 Strategy outlines their main areas of focus, including key topics such as Copyright Reform, Digital Humanities, Open Access, Metrics, and Research Data Management.

Funding	Funded by institutional library membership
Comments	Repositories are only part of a much broader remit for this group Important advocacy group

Name	National Institute of Informatics (NII) https://www.nii.ac.jp/ https://www.nii.ac.jp/en/
Description	The NII is an inter-university Japanese research institute created in April 2000 for advancing the study of informatics. The mission of NII is to create future value in informatics. Research at NII focuses on information-gathering techniques and systems for information management. The NII in its push for open access and open science drives the development of academic information infrastructures. The NII oversees and maintains a searchable information database called Webcat and Webcat Plus.
Software (note open or closed)	Webcat is an open access database that supplies holdings information for materials held in research institutes and university library collections throughout Japan.
Metadata requirements	Not specified
Recommendations for interoperability	Not specified
Funder requirements	Not specified
Policy requirements	Not specified

Key relationships	SPARC Japan
Compliance with FAIR principles	Not specified
Support available/ Could be person	soumu@nii.ac.jp Kazu Yamaji, National Institute of Informatics
Notes	Regional update from Japan's NII in April 2017

Name	SPARC
Description	SPARC (the Scholarly Publishing and Academic Resources Coalition) is a US based coalition. Close relationships with international affiliate organizations SPARC Europe, SPARC Japan, and the newly launched SPARC Africa. Also relationship with COAR.
Funding	Funded by institutional library membership and grants.
Comments	Important advocacy group.

Name	UKCORR United Kingdom Council of Research Repositories http://ukcorr.org/
Description	The professional organisation for UK open access repository administrators and managers. **Vision** UKCoRR has a vision of the work of repository management as a professionally recognised and supported role within UK research institutions. **Mission** UKCoRR is an independent body for repository managers, administrators and staff in the UK that:

	• Promotes repository management as a recognised and respected profession • Provides a forum for discussion and exchange of experience • Represents the views and concerns of those who work with repositories in organisational, policy and strategic development **UKCoRR and research** UKCoRR has no funding to act as a research or development body. Where appropriate, issues identified by the membership will be passed on to appropriate bodies on behalf of UKCoRR. *(From here)*
Software (note open or closed)	N/A however UKCoRR resources are open to the public at https://drive.google.com/drive/folders/0B93C0Wu5SHkbQzNXUjVyc1FNb1k
Metadata requirements	N/A
Recommendations for interoperability	N/A
Funder requirements	N/A
Policy requirements	UKCoRR has no funding to act as a research or development body. Where appropriate, issues identified by the membership will be passed on to appropriate bodies on behalf of UKCoRR.
Key relationships	The professional organisation for UK open access repository administrators and managers.
Compliance with FAIR principles	N/A
Support available/ Could be person	**UKCoRR Technical Officer** The role of Technical Officer requires a good knowledge of repository software and Current Research Information Systems, relevant protocols and standards (e.g. OAI-PMH, SWORD, RIOXX etc.), along with an understanding of the broader scholarly communications ecosystem and relevant service providers (e.g. ORCID, Jisc Publication Router, CORE, Europe PMC, IRUS-UK etc.)

	The role involves: Maintaining the visibility of UKCoRR as a source of technical expertise in the sector, by: ● acting as a point of contact for technical aspects of repository development for the UKCoRR membership and wider community ● commenting on technical issues as they relate to repositories, open access and scholarly communication via the mailing list, UKCoRR blog and other suitable outlets ● contributing to members' events from a technical perspective ● liaising with other key stakeholders on technical developments, notably Jisc, and contributing to their consultation and outreach events
Notes	The professional organisation for UK open access repository administrators and managers. May be useful to get in touch with the UKCoRR Technical Officer for the scope of this CAUL project

Appendix 4: Force 11 FAIR principles
#3: Repository User Stories

TO BE FINDABLE:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.

F2. data are described with rich metadata.

F3. (meta)data are registered or indexed in a searchable resource.

F4. metadata specify the data identifier.

TO BE ACCESSIBLE:

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.

A1.1 the protocol is open, free, and universally implementable.

A1.2 the protocol allows for an authentication and authorization procedure, where necessary.

A2 metadata are accessible, even when the data are no longer available.

TO BE INTEROPERABLE:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles.

I3. (meta)data include qualified references to other (meta)data.

TO BE RE-USABLE:

R1. meta(data) have a plurality of accurate and relevant attributes.

R1.1. (meta)data are released with a clear and accessible data usage license.

R1.2. (meta)data are associated with their provenance.

R1.3. (meta)data meet domain-relevant community standards.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Appendix 5: Functional areas for repository user stories #3: Repository User Stories

Discovery/search

UX/help/guidelines

Collection management

Access - download/print/save

Embargo management

Rights management

Permission management

Preservation

PIDs/citations

Integrations/linking

Research lifecycle

Notifications

Engagement - editing/tagging/requests

Harvesting - content transfer and movement

Batch ingestion

Versioning/event tracking/auditing

Remixing and re-use

Impact and ROI

Reporting

Configurable metadata

SLA/terms

NTROs

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Appendix 6: Full list of recommendations identified from Products #1-#3

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

The Product #2 report presented four recommendations:

1. That CAUL seek specific project funding to develop required national/regional repository infrastructure.
2. That CAUL seek a consortial membership of COAR.
3. That CAUL set up an ongoing repository technical advisory working group.
4. That CAUL set up a group which reviews the training and professional development required for repository staff.

Overall:

- Agreement with the 4 recommendations with initial priority to recommendation 3: that CAUL set up a technical advisory group as key to getting consensus and expert advice on the minimum metadata standards and a repository checklist.
- CAUL technical advisory working group to provide advice and recommendations on how to participate in technical initiatives like Scholix and re3data.org.
- CAUL to provide funding for coordination and planning on issues.
- Adoption of an Institutional Open Access Policy.
- Legal advice on what an Open Access Policy should say.
- Advice from CAUL on recommended minimum requirements for repository systems (including next generation repository systems). Possible changes to copyright act – changes to access gains importance including access conditions.
- Standardised reporting across the sector for compliance as a way to measure increase in open access usage. This is proposed along the lines of the IRUS-UK model that provides COUNTER-conformant usage statistics for all content downloaded from participating UK institutional repositories.
- Development of a recommended preservation strategy for all resource types.

Findability:

- Tags to differentiate between green open access, gold open access, and free to read open access.
- Agreed core metadata and minimum metadata requirements for Australasian university libraries.
- Ensure that metadata is extensible to align to standards that emerge.
- Tags to track funder compliance.

- Recommend ORCiD ID and ORCiD Ringgold identifier for institutions as an integral tool for disambiguation.
- A field for funder - either compliance or acknowledgement.
- Need to integrate seamlessly into the researcher profile.
- Repositories should disseminate metadata widely to aggregators (UnPaywall, BASE etc.). Use persistent identifiers in metadata where possible.
- An Australian standardised metadata
- Agreed way to deal with ORCID iD (as a name authority tool or just another piece of metadata in the record).

Accessibility:

- Repositories should be as open as possible and as closed as necessary.
- Repositories should use open (free; non-proprietary) and standardised communication protocols to enable seamless access to anyone with a computer and the Internet (no need for specialised tools to gain access to content).
- Protocols must comply with Web Content Accessibility Guidelines (WCAG) so that repository content can be meaningfully accessed and navigated by people with disabilities.
- Systems and workflows should automatically classify and protect sensitive data. Persistence and preservation of research output and other digital assets should be a research and administrative priority. Datasets disappear over time with implications not just for access but for provenance.

Interoperability:

- CAUL technical advisory group to investigate moving from OAI-PMH to ResourceSync.
- Repositories should link data, publications, and grant information as much as possible
- Adoption of licensing standards in machine readable formats.
- Wider and uniform participation in harvesting agencies with CAUL providing a recommended list of universal targets e.g. TROVE, UnPaywall and discipline specific harvesters.
- Each repository should expose/share/publish their metadata schema and any crosswalks that they have created for publishing to and harvesting from, systems such as TROVE or dataCITE etc.

Reusability:

- Standardised field for a creative commons licence.
- Inclusion of a tag for open access and availability of a licence.
- Be able to attach clear and accessible data usage licenses, such as Creative Commons (CC) licenses, in forms that are both human- and machine-readable.
- Tags to identify the version to facilitate reusability.
- Provenance metadata to enable appropriate attribution and citation and to help researchers to determine whether they trust the data (for re-use).

Appendix 7: Non-functional requirements

#6 Next generational repository tools and general requirements

	Non-Functional Requirements	***Priority***	***Comments***
	Service Provision and performance		
A1.01	The solution is hosted		
A1.02	The solution is available to end-users for a minimum of 99.5% of the time, excluding predefined maintenance schedules		
A1.03	The solution is scalable to meet the load of large numbers of users without impact on performance		
A1.04	The solution has a static IP range to enable trusted communications with solution receiver		
A1.05	The service receiver is satisfied as to the size capability of the solution (e.g. able to upload files up to 20 TB)		
A1.06	The service receiver is satisfied as to the records capability of the solution (i.e. any limits as to number of records without impact on system performance)		
A1.07	The solution provides future-proofing for new file formats (e.g. 3D imagery and virtual reality content)		
	Exit Strategy		
A2.01	The solution provider will delete the service receiver's hosted data upon request from the service receiver		
A2.02	The service receiver's hosted data is downloadable for the entirety of the contract period in a		

	format agreed to by the service receiver		
	Content ownership		
A3.01	The service receiver retains ownership of all hosted data added to or created within the solution		
A3.02	The service receiver understands and is in agreement with any use made of the service receiver's hosted data by the solution and/or solution provider		
A3.03	The service receiver understands and is in agreement with the use and purpose of metadata and statistical information generated as to solution usage		
A3.04	The service receiver is satisfied with the solution providers management in relation to data ownership as to the service receivers data where the solution provider uses an upstream provider of network or storage		
	Compliance		
A4.01	Solution complies with Australian Privacy Act 1988; specifically to: Part IIIC (Notification of eligible data breaches), Part VIII (Obligations of confidence), Part IX (Miscellaneous, 95: Medical research, 95A: health information, 95AA genetic information) and Schedule 1 (Australian Privacy Principles)		
A4.02	The service receiver understands and is satisfied with the solution provider's policy and notification procedures should law enforcement agencies request access to hosted data		
A4.03	The solution is able to comply with the State Archives act		

	Data Centre location		
A5.01	The service receiver is satisfied with the geographical location/s of the data centre/s where the service receiver's data will be kept by the solution provider		
A5.02	The service receiver is satisfied as to the geographical location/s of the solution's backup/disaster recovery systems; i.e. ability to replicate repository metadata, linkages and streams to enable disaster recovery instance		
A5.03	The service receiver understands and is satisfied with the level of physical security in place for the solution's Data Centre/s		
	Data Centre security protocols		
A6.01	The service receiver is satisfied with the solution provider's policy and procedures around information security in relation to solution provider's suppliers		
A6.02	The service receiver is satisfied with the solution provider's incident response and management plans in place to minimise the impact of unauthorised disclosure		
A6.03	The service receiver is satisfied with the solution provider's password policy		
A6.04	The service receiver is satisfied with the documented access control policy at the solution's Data Centre/s		
A6.05	The solution provider has a process in place for detecting information security vulnerabilities		
A6.06	The service receiver is satisfied with the solution provider's ToS and SLA which set out compensation clauses for breaches caused by vulnerabilities in the service		
A6.07	The service receiver is satisfied with the solution provider's policy in relation to removeable media with the solution's Data Centre/s		
A6.08	The service receiver understands and is satisfied with the solution provider's management of staffing roles and accountability to reduce opportunities for unauthorised modification or		

	misuse of information		
A6.09	The solution provider restricts and monitors privileged utility programs related to the solution		
A6.10	The service receiver understands and is satisfied with the background verification check policy of the solution provider in relation to new and existing staff		
A6.11	The solution provider requires employees, contactors and third party workers to sign confidentiality and non-disclosure agreements		
A6.12	The service receiver is satisfied with the level of security awareness training received by security provider employees, contractors and third party workers		
A6.13	The service receiver is satisfied with the solution provider's incident response and management plans to system failure and recovery		
A6.14	The service receiver is satisfied as to how the solution provider guarantees continuity of access to its services in the event of an outage		
A6.15	The service receiver is satisfied and in agreement with the contract limits and provision for insurance, liability and indemnity for incidents where data has been lost or compromised		
	Data Integrity and Security		
A7.01	The service receiver is satisfied and in agreement with the documented policy, roles, responsibilities and notification procedures of the solution provider and the service receiver during and after an incident where data has been lost or compromised		
A7.02	The solution has greater than a single point of failure		
A7.03	The solution has a scheduled backup regime which meets the service receiver's requirements		
A7.04	The solution validates backups to confirm efficacy		

A7.05	The solution provider undertakes technical compliance reviews of the solution at a frequency satisfactory to the service receiver		
A7.06	The solution provider maintains an information security programme that complies with COBIT, ISO 27000, NIST 800 Series and ISAE 3402		
A7.07	The solution provider has an annual independent security audit from a qualified security auditor		
A7.08	The solution provider has CSA STAR Certification and/or Attestation.		
A7.09	The solution provider allows for post-deployment penetration tests (and other similar tests) conducted by the service receiver		
A7.10	The solution provider has a Security Information Event Monitoring (SIEM) service that logs and monitors all logical access to data which identifies individuals		
A7.11	The solution's event logs are maintained and undergo scheduled review by solution provider		
A7.12	The solution provider has firewalls to protect against network threats		
A7.13	The solution provider uses intrusion detection systems (IDS) and/or intrusion protection systems (IPS)		
A7.14	The solution provider or the solution uses malware detection for system and software including a process to control the installation of software onto operational systems		
A7.15	Data classified by the service receiver as 'sensitive' and above is stored in Australia		
A7.16	The service receiver's personal data, if stored by the solution provider, is stored in Australia		
A7.17	The solution identifies and discounts bad robot activity from repository usage statistics		
	Support		

A8.01	The solution provider's ToS and SLA describe how confidentiality, integrity and availability of official information and the privacy of all personally identifiable information is protected		
A8.02	The service receiver understands and is satisfied with the warranty period for the solution		
A8.03	The service receiver is satisfied and in agreement as to the policies and processes relating to solution upgrade and fixes		
A8.04	The service receiver is satisfied and in agreement as to how software upgrades are documented in the SLA		
A8.05	The solution has a robust user community working closely with the solution provider		
A8.06	The solution provider releases a roadmap as to developmental direction for the solution and openly shares the roadmap with customers for feedback		
A8.07	The solution provider hosts and releases plugins and scripts to the user community		
A8.08	The solution provides a staging/testing environment that is separated from the operational (production) environment		
A8.09	The service receiver has access to a local instance staging/testing environment that mirrors the operational environment		
A8.10	The solution provides help manuals, community, forums, responsive help desk and FAQ		
A8.11	The solution provides advanced notification as to scheduled and unscheduled outages		
A8.12	The solution offers an automated regression testing regime after installation of an upgrade		
	Branding		
A9.01	The solution provides a unique URL, of the service receiver's choice, to the repository instance		

	Preservation		
A10.01	The service receiver understands and is satisfied with the preservation capabilities of the solution		
A10.02	The solution complies with digital preservation standards (e.g. ISO 14721:2012)		

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

	End User Experience Requirements	Priority	Comments
	User Interface (UI)		
B1.01	SSL connection is used throughout the solution		
B1.02	The solution complies with WCAG 2.1 standard to AA level		
B1.03	The solution is accessible via internet connection from anywhere		
B1.04	The solution is usable to the base standard for desktop browsers		
B1.05	The solution is mobile responsive		
B1.06	The solution is fully customisable; i.e. able to provide a customisable interface across the repository and for specific collections (including; local branding and control of style, images and graphical elements) without the need for programming skills		
B1.07	The presentation of the interface, records and digital items is intuitive and attractive		
B1.08	The solution delivers streamlined navigation across a range of devices		
B1.09	The solution delivers streamlined navigation between digital items within repository collections		
B1.1	The solution delivers streamlined navigation between digital items within a result set		

0			
B1.11	The solution provides the ability to link to 'help' resources within the application		
	Search		
B2.01	The repository content is indexed in real time		
B2.02	The search function is intuitive and easy to use		
B2.03	The search function delivers a fast and reliable response time		
B2.04	The search function offers both basic (keyword) and Advanced (multiple field) search		
B2.05	The search function can deliver results from the entire repository in one query		
B2.06	The search function can deliver results from a selection of repository collections in one query		
B2.07	The search function can deliver results from within a single repository collection in one query		
B2.08	The search function enables results to be limited to a subset of repository content or a collection		
B2.09	The search function can be conducted across all digital item format types in one query		
B2.1	The search function enables filter/refinement of large results sets through facets or pre-		

0	search filters		
B2.11	The Advanced search function provides the option to search words or phrases across multiple fields simultaneously		
B2.12	The search fields are configurable for queries; i.e. search fields, including custom fields, can be added or removed from search function		
B2.13	The search fields are configurable for display in the search results		
B2.14	The search function delivers highlighted search term/s in full text and/or search results		
B2.15	The search function allows authenticated users to save and manage searches and search result sets across user sessions		
B2.16	The search function enables interoperability with third-party content providers (e.g. Cited by, Citation)		
B2.17	The search function returns results from full text digital items across the repository		
B2.18	The search function returns results from full text digital items across a selection of repository collections		
B2.19	The search function returns results from full text digital items within a repository collection		
B2.20	The search function provides the ability to browse/search by repository collection, author and organisational affiliation		
B2.21	The search function delivers suggestions/recommendations in line with search; e.g. "You liked this, you may also like that"		

	Accessibility		
B3.01	The solution interoperates with assistive software (JAWS, Dragon, etc.)		
	Content engagement		
B4.01	The solution displays links to the content that are prominent in the search result list and the record display		
B4.02	The solution displays digital items in a user friendly format		
B4.03	The solution makes use of multiple viewers related to the type of digital item being viewed (e.g. pdf reader, audio player, video player, page-turning book interface, etc)		
B4.04	The solution links to and/or provides human readable information about licence conditions		
B4.05	The solution displays the file size of the digital item		
B4.06	The solution clearly identifies/labels Open Access content		
B4.07	The solution displays embargo period details (where metadata only for digital item is publicly available)		
B4.08	The solution provides download of individual digital items via a range of methods (e.g. direct download, email, export to, etc.)		
B4.09	The solution provides download of individual records and groups of records via a range of methods (e.g. email, export to …., etc.)		
B4.1	The solution enables export to citation and referencing software (e.g. EndNote, RefWorks,		

0	etc.)		
B4.11	The solution provides social media sharing, commenting and tagging (e.g. Activity Streams 2.0, Web Annotation Model and Web Annotation Protocol)		
B4.12	The solution provides creation of record 'Favourites' retained across user sessions for authenticated users		
B4.13	The solution provides creation of record 'Favourites', session dependant, for non-authenticated users		
B4.14	The solution allows authenticated users to create individual collections/sets sourced from existing repository collections		
B4.15	The solution is able to generate RSS feeds for newly available digital items		
B4.16	The solution is able to showcase 'new' content		
B4.17	The solution allows related content to be discovered as a grouping (package)		
B4.18	The solution is able to stream multimedia types		
B4.19	The solution enables the creation of online gallery/exhibitions/themes		
B4.20	The solution is able to lead the user through a curated online exhibition		
B4.21	The solution supports text/data mining		

B4.22	The solution is able to display a 'take down' notice (with an underlying workflow) at digital item level		
	Content access		
B5.01	The solution is able to clearly display access rights for each digital item (e.g. able to view in repository but unable to download)		
B5.02	The solution is able to display a 'Request content' button, with the ability to create a workflow underneath, for content with restricted access		
	Interoperability		
B6.01	The solution is able to link to digital items outside the repository		
B6.02	The solution is able to link to other repositories		
	Reporting		
B7.02	The solution provides metrics in an attractive and informative way (Downloads, citations, views, etc.)		
B7.03	The solution provides metrics on impact and social media engagement (e.g. journal and data citation and alternative metrics)		

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

	Repository Management Requirements	Priority	Comments
	Administrative User interface (UI)		
C1.01	A SSL connection is used throughout the solution		
C1.02	The administrative interface complies with WCAG 2.1 standard to AA level		
C1.03	The administrative interface is usable to the base standard for desktop browsers		
	Collection development (ingest/deposit)		
C2.01	The solution offers a range of options to ingest/deposit content		
C2.02	The solution provides ingest of digital items via SWORD or similar deposit protocols		
C2.03	The solution provides deposit via a web interface		
C2.04	The deposit workflow provides clear form fields that are easy to populate		
C2.05	The solution provides the ability to drag and drop into the system		
C2.06	The solution provides a variety of locally customisable web forms for deposit of different content types		
C2.07	The solution provides depositors with standardised lists of terms to describe their submission (e.g. FOR codes and resource types)		
C2.08	The solution is able to clearly identify suggested resource types as part of the deposit workflow		
C2.09	The solution is able to deliver/present an interactive checklist to facilitate the deposit, including compliance, workflow		

C2.10	The deposit workflow provides instructions and links to instructional documentation associated with self-deposit at point of need		
C2.11	The solution offers scheduling options for harvest activities		
C2.12	The solution allows bulk ingest of digital item/data stream with accompanying metadata		
C2.13	The solution allows bulk ingest of metadata		
C2.14	The solution allows bulk ingest of metadata from a delimited file or spreadsheet		
C2.15	The solution allows ingest of multiple file types in a multipart ingest		
C2.16	The solution is able to upload and manage a range of formats (e.g. PDF, doc, docx, xls, xlsx, jpeg, jpg, png, gif, tiff, mp3, mp4, flv, wmv, avi, mov, aiff, wav, wma, m4a)		
C2.17	The solution is able to ingest and manage the code used to analyse data		
C2.18	Digital items receive virus scanning during ingest/deposit		
C2.19	The solution provides validation of digital items during ingest/deposit		
C2.20	The solution identifies and reports on record duplication at ingest/deposit		
C2.21	The solution automatically generates metadata from ingested files		
C2.22	The solution extracts technical data from ingested/deposited digital items (where present)		
C2.23	The solution provides the ability for the author/creator to consent to deposit their work		
C2.24	The solution provides the ability for the author/creator to assert copyright ownership/permission for deposited digital items		
C2.25	The solution provides the ability for the author/creator to confirm compliance with Privacy Laws and other legal provisions for deposited digital items		
C2.26	The solution provides the ability for author/creator to delegate deposit activity to a 3rd party		

C2.27	The solution is able to ingest records using a variety of common metadata schemas, and is extensible to additional formats		
C2.28	The solution is able to ingest a range of file sizes and is able to set a maximum file size for deposit		
C2.29	The solution is able to index the full text of text based digital items at ingest		
C2.30	The solution is able to perform OCR on text based digital items at ingest		
	Content management		
C3.01	The solution accommodates workflow stages; input, edit, approve		
C3.02	Workflows can be defined and customised locally		
C3.03	The solution provides the ability to update information within records (e.g. metadata, links and digital items) via a web interface or client		
C3.04	The solution provides online (in browser) help/prompts for data entry fields		
C3.05	The solution provides the ability to edit all records via an online editor (including any element, field, subfield, or fixed value field value as appropriate for the format)		
C3.06	The solution allows the export of digital items		
C3.07	The solution allows the export of metadata		
C3.08	The solution allows the bulk export of digital items		
C3.09	The solution allows the bulk export of metadata		
C3.10	The solution allows deletion of digital items		
C3.11	The solution allows deletion of metadata records		

C3.12	The solution supports the display and recording of more than one identifier (e.g. Grant, Author ID, Publication ID, etc.)		
C3.13	The solution supports annotations to non-archival (access) copies of digital items		
C3.14	The solution provides the ability to establish a repository collection/sub collection hierarchy		
C3.15	The solution provides the creation and management of unique repository collections and sub collections based on identified criterial (e.g. topic and/or 'Dark Archive' - with restricted access files)		
C3.16	The solution supports the recording of provenance		
C3.17	The solution provides the ability to reconcile duplicate records and provide notification of duplication		
C3.18	The solution is able to incorporate and link supporting data (enabling the ability to render the evolution of the work)		
C3.19	The solution is able to comply with 'Record Keeping system' requirements		
C3.20	The solution provides an audit trail reporting on all changes to data within the solution		
C3.21	The solution provides versioning, including the ability to view and roll back to past versions of an access copy of a digital item and metadata		
C3.22	The solution enables the use of internal linked data that can be extended between digital items in the repository (e.g. "Works by this author in our collection")		
C3.23	The solution is able to automatically generate a coversheet from specified metadata fields that clearly identify the source of the digital item, citation and conditions of use, and integrate with the digital item		
C3.24	The solution provides the ability to view the metadata via a browser by adding .rdf to URL		
C3.25	The solution is able to create 'access' format and 'preservation' format for each digital item		

C3.26	The solution is able to incorporate, display, store and report on a broad range of identifiers (e.g. Grant ID, PubMed ID, ORCID, Social Network Identities, WebID, etc.)		
C3.27	The solution allows digital items to be placed in more than one collection		
C3.28	The solution allows locally defined embargo periods per digital item		
C3.29	The solution provides automatic release for digital items when embargo period expires		
	Content control		
C4.01	The solution provides conversion of TEI XML content to HTML		
C4.02	The solution provides the management of TEI XML [and related formats]		
C4.03	The solution provides conversion of digital items to ePub and other e-reader friendly formats		
C4.04	The solution provides link validation, and link resolution testing functionality		
C4.05	Text in all records is Unicode 6.0 compliant for importing, editing, storage and export		
C4.06	Solution will support Unicode 8.0 in the future		
C4.07	The solution allows for the creation of single/multiple part digital item(s)/data stream(s)		
C4.08	The solution uses only approved encryption protocols and algorithms		
C4.09	The solution provides the creation of new access copies from deposited copies		
C4.10	The solution provides the generation of access copy derivatives, in different formats, from a deposited digital item		
C4.11	The solution supports characters or glyphs which do not qualify for inclusion in Unicode standards		
C4.12	It is possible to import and export from/to the solution without loss of data		

C4.13	The solution supports flexibility with filenames		
C4.14	The solution encrypts the digital item at read and during transmission		
C4.15	The solution is able to generate HTTP links with appropriate link relation types and format indicators to interlink (digital item, metadata, licence) the web resources that make up the repository record (i.e. Signposting)		
	Metadata management		
C5.01	The solution indexes all metadata fields		
C5.02	The solution enables simple and efficient management of metadata		
C5.03	The solution enables the ability to create local, indexable and reportable, metadata fields		
C5.04	The solution allows for the repair and/adjustment of metadata		
C5.05	Metadata records can be updated/changed in bulk (e.g. add, modify, insert, delete, supress)		
C5.06	The solution provides the same editing capabilities for all metadata formats		
C5.07	The solution supports a variety of standards-based metadata schemas (e.g. simplified Dublin Core, qualified Dublin Core, METS, MODS, PREMIS, TEIXML, RIF-CS, OAI-ORE and MARCXML)		
C5.08	The solution reports on metadata validation results		
C5.09	Metadata can be entered in multiple languages (including appropriate diacritics)		
C5.10	Repository administrators are able to add and change controlled vocabulary fields		
C5.11	The solution supports controlled lists against some metadata fields; either held locally or drawn from external sources (e.g. publisher site, DOI, etc.)		
C5.12	The solution provides the ability to cross walk the metadata		

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

C5.13	The solution supports association of single or multiple files with one metadata record		
C5.14	The solution supports RDF compliance for metadata		
	Rights Management		
C6.01	The solution is able to attach rights management information to individual digital items (e.g. Licence agreements, copyright statements, usage guidelines, cultural sensitivity warning, etc.)		
C6.02	The solution supports Creative Commons licence types		
C6.03	The solution stores author/creator consent (compliance statement) for future retrieval		
C6.04	Publisher permissions can be stored in the solution, integrated with metadata records and viewable by repository administrators		
C6.05	The solution is able to embed a HTTP link to point to the URI of the licence applied to a digital item (i.e. Signposting)		
C6.06	The solution supports ODRL and similar machine readable rights languages		
	End User access management		
C7.01	The solution provides the ability to review ingested/deposited digital items prior to release to public view (mediation)		
C7.02	The solution provides the ability to assign tiered or restricted client access to repository collections		
C7.03	The solution provides tiered or restricted client access related to digital item licences (e.g. Creative Commons and other copyright licences)		
C7.04	The solution provides the ability to assign tiered or restricted client access to individual digital items		

C7.05	The solution provides the ability to control digital item access; i.e. download, view/play/listen on site only access, limit to copy and paste function (image application), lower resolution for downloaded imagery		
C7.06	The solution enables restricted access based on IP range to repository collections and/or digital items		
	Administrative User interface access management		
C8.01	The solution enables tiered access; i.e. different user roles and permission, and controls different levels of access by user group to the Administrative Interface		
C8.02	Repository administrators are able to assign/reassign requests to other repository administrators for processing		
C8.03	The solution enables restricted access based on IP range (i.e. on campus access)		
C8.04	The solution provides for an administrative role with read only access		
	Content delivery		
C9.01	Individual records and/or digital items can be supressed from display		
C9.02	The solution is able to link a metadata record to content outside the repository		
C9.03	The solution is able to generate and store system generated thumbnails and delivery renditions of high quality images, audio and video		
C9.04	The solution allows the generation of a 'suggested' or 'selected' citation format for metadata records and associated digital items		
	Preservation		
C10.01	The solution maintains data in a preservation-ready state		

C10.02	The solution incorporates the storage and preservation of metadata and digital items		
C10.03	The solution is able to migrate to new and emerging data formats overtime		
C10.04	The solution conforms to OAIS standard and reference model		
C10.05	The solution provides fixity check function which can be schedule over all, or identified parts, of the repository		
C10.06	The solution provides a check sum function which can be scheduled over all, or identified parts, of the repository		

	Functional System Requirements	*Priority*	*Comments*
	Reporting		
D1.01	Use of the reporting function does not impact on application performance		
D1.02	The solution provides a reporting dashboard		
D1.03	The solution allows the interrogation and analysis of repository collection and sub collection/s content within the solution		
D1.04	The solution is able to report on content in individual repository collections		
D1.05	The solution records and reports on the number of End User interface hits, metadata record views, downloads, media views (inside the solution), demographic information, top items, top authors, etc.		
D1.06	The solution provides reports as to high volume demand on individual digital items and/or repository collections		
D1.07	The solution provides operational and usage reports for repository administration activities		
D1.08	The solution enables the creation of repository administrator defined reports that are easily created, modified, saved and scheduled using an intuitive user interface, without the need for specialised database query skills		
D1.09	The solution outputs reports in a variety of formats (e.g. PDF, Excel, CSV)		
D1.10	The solution allows for the identification and tracking of deposit (so that depositors can be targeted for follow up)		
D1.11	Solution reports can be configured to be received by specified email addresses		

D1.12	The solution is able to generate reports based on funder reporting requirements		
D1.13	The solution is able to report on OA compliance within the repository		
D1.14	The solution provides the ability to track every digital item that comes into the repository and discover the source of the individual digital item		
D1.15	The solution is able to report on digital item usage throughout the digital item lifecycle		
D1.16	The solution is COUNTER, SUSHI, etc. compliant		
	Notifications		
D2.01	The solution provides automated, customisable email notifications for a variety of use cases; e.g. new record alert, successful deposit, usage notification (individual and faculty level) and notifications to be sent to repository administrators, creators and end users		
	Capabilities		
D3.01	The solution has been built around open standards		
D3.02	The solution API is standards based, documented and supported		
D3.03	The solution meets RESTful criteria		
D3.04	The solution supports the creation and use of 'Cool' URLs (https://www.w3.org/Provider/Style/URI)		
D3.05	The solution is compatible with linked open data		
D3.06	The solution is able to export JSON data		
D3.07	The solution supports IIIF		
D3.08	The solution supports IPFS		

D3.09	The solution supports ResourceSync and ResourceSync Change Notifications		
D3.10	The solution complies with Linked Data Platform rules		
D3.11	The solution supports Webmention		
D3.12	The solution supports WebSub		
D3.13	The solution supports messaging protocols (e.g. AMQP, Kafka)		
D3.14	The solution supports encryption key escrow		
	Interoperability		
D4.01	The solution is able to authenticate against the service receiver's organisational authentication systems		
D4.02	The solution accommodates a feed from the service receiver's HR and student management systems		
D4.03	The solution enables automated data exchange (digital item and metadata) with research systems in relation to publications (e.g. push to/pull from CRIS system)		
D4.04	The solution accommodates a feed from the service receiver's Grant system		
D4.05	The solution accommodates a feed from the service receiver's Ethics system		
D4.06	The solution provided web forms for deposit can be integrated with local systems for prepopulation of fields to look up on names, affiliations, grants, subject/FOR codes, etc.		
D4.07	The solution enables integration with the service receiver's records management system		
D4.08	The solution is able to push data to the service receiver's researcher profiles		
D4.09	The solution integrates with existing handle server (handles to metadata and digital items) including management and maintenance		

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

D4.10	The solution is able to integrate with the ANDS, and other, DOI Minting Services		
D4.11	The solution interoperates (push to and pull from) with the solution receiver's archival research data storage		
D4.12	The solution enables migration of metadata to and from the solution receiver's LRMS		
D4.13	The solution is able to interoperate with EZproxy functionality (if tiered access is unable to be achieved within the solution)		
D4.14	The solution is able to integrate with other Discovery platforms (e.g. Primo, Summon)		
D4.15	The solution interoperates with Australian Access Federation (AAF) authentication (for sharing of research outputs)		
D4.16	The solution is fully compliant with OAI feeds for interoperability with diverse systems		
D4.17	The solution is able to integrate with Research Data Australia (RDA)		
D4.18	The solution supports integration with Data Citation Index		
D4.19	The solution accommodates lookup authority services within metadata input (i.e. ORCID, etc)		
D4.20	The solution enables integration with publishers; e.g. accepted post print articles automatically sent from publisher to repository		
D4.21	The solution provides an automated process to scan for post prints and datasets (with indexed metadata) and then proceed to automatic ingest into the repository		
D4.22	The solution integrates with external digital preservation tools		
D4.23	The solution provides an API to export metadata to third party providers (e.g. Sherpa/Romeo)		
D4.24	The solution integrates with Sherpa/Romeo for OA checking (Copyright)		

D4.25	The solution allows integration with third party analytics tools (e.g. Tableau and Power BI)		
D4.26	The solution enables integration with Google Analytics Solutions, including, but not limited to Google Tag Manager, Google Search Console Clicks and Google Analytics Events		
D4.27	The solution integrates with IGSN (New Sample ID) minting service		
D4.28	The solution integrates with CHOR-AUS		
D4.29	The solution is suitable for incorporation into RAMP		
D4.30	The solution is able to integrate with research networks (e.g. HuNI)		
D4.31	The solution is able to integrate with ResearchGraph.org		
D4.32	The solution supports Linked Data Notifications		
	Discoverability		
D5.01	The solution embeds machine-readable metadata (read metatags) that is independent of UI customisation		
D5.02	The solution is able to produce a date stamped (date published and/or date added) sitemap		
D5.03	The solution is indexed by Google Scholar, Trove, Google, OAIster, Bing and Microsoft Academic		
D5.04	The solution is harvested, indexed and searchable by a range of search engines/tools including full text indexing of text based digital items (e.g. CORE, BASE, OpenAIRE, Trove)		
D5.05	The solution allows harvest from a discrete repository collection (i.e. discrete sets)		
D5.06	The solution provides the ability to embed search boxes on external web pages		
D5.07	The solution is able to create customised widgets that can be embedded on another website (e.g. latest additions showcased at faculty website, author badges to be embedded in		

	personal blogs or researcher profiles).		
D5.08	The solution is able to generate persistent identifiers (HTTP(S) URIs) for all digital items		
	Migration		
D6.01	As part of the legacy migration process; relationships between digital item and metadata is retained (including handles and access restrictions)		
D6.02	The solution is able to ingest existing digital items and associated metadata into one or more repository instance (as part of the legacy migration process)		
D6.03	The solution is able to maintain existing repository ID numbers as part of the legacy migration process		
D6.04	The integrity of existing links and metadata is maintained post migration		

Appendix 8: Feedback provided at the CAUL 2018 / 2 meeting (September)

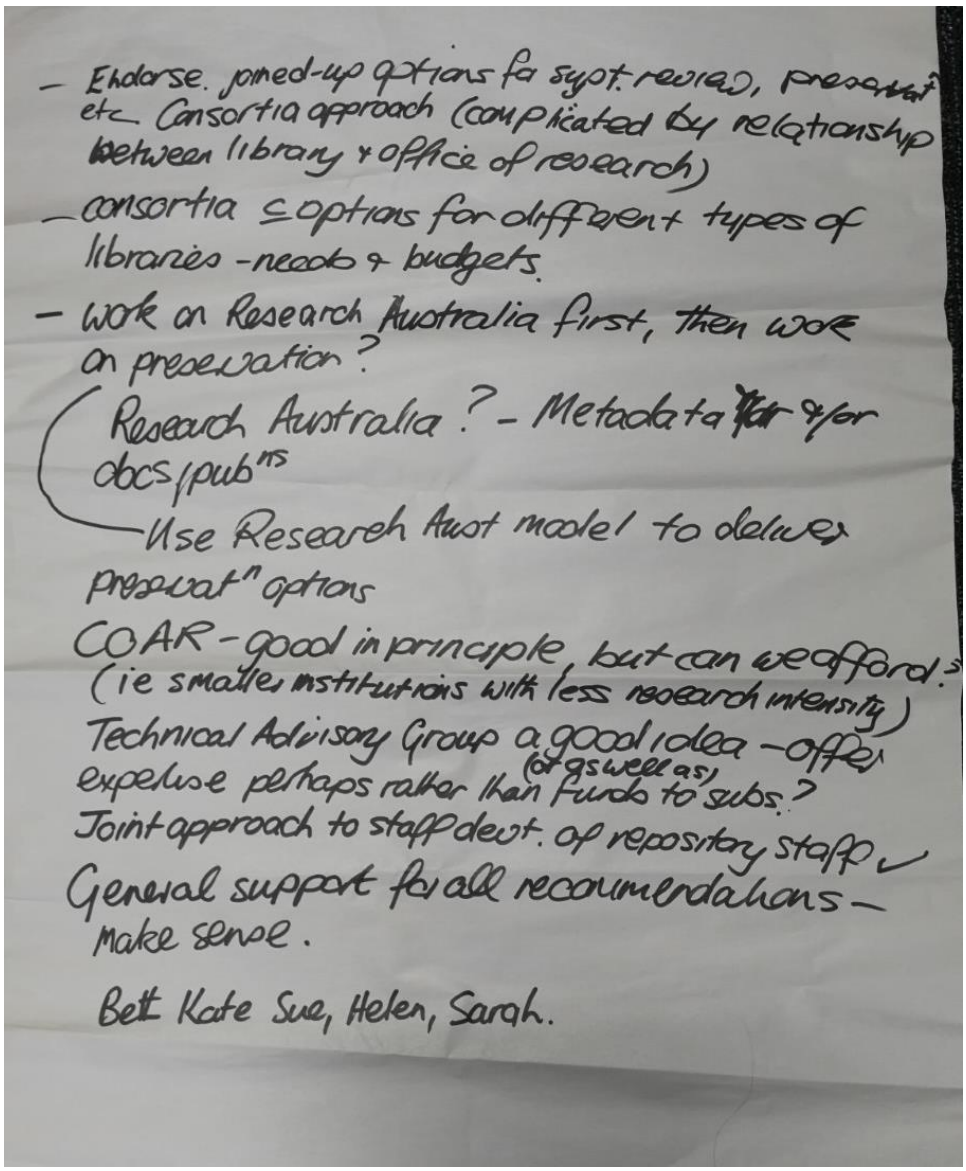
Review of Project Products

At the CAUL Council 2018/2 (September) meeting, attendees were asked to provide feedback to the draft project report using butcher paper notes. Feedback is provided here as Appendix 8.

A secondary feedback process was undertaken in February – March 2019 via a Zoom videoconference and online survey. Feedback from this process is provided in Appendix 9.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

By Bett, Kate, Sue, Helen, and Sarah



- Endorse, joined-up options for support review, presentation etc Consortia approach (complicated by relationship between library & office of research)
- Consortia c options for different types of libraries – needs & budgets
- Work on Research Australia first, then work on presentation?
  ->Use Research Australia model to deliver present options

Research Australia? – Metadata for dpcs/pubs

COAR – good in principle, but can we afford?

(ie. smaller institutions with less research intensity)

Technical Advisory Group a good idea – Offer expertise perhaps rather than (or as well as) fur to subs?

Joint approach to staff dev. of repository staff

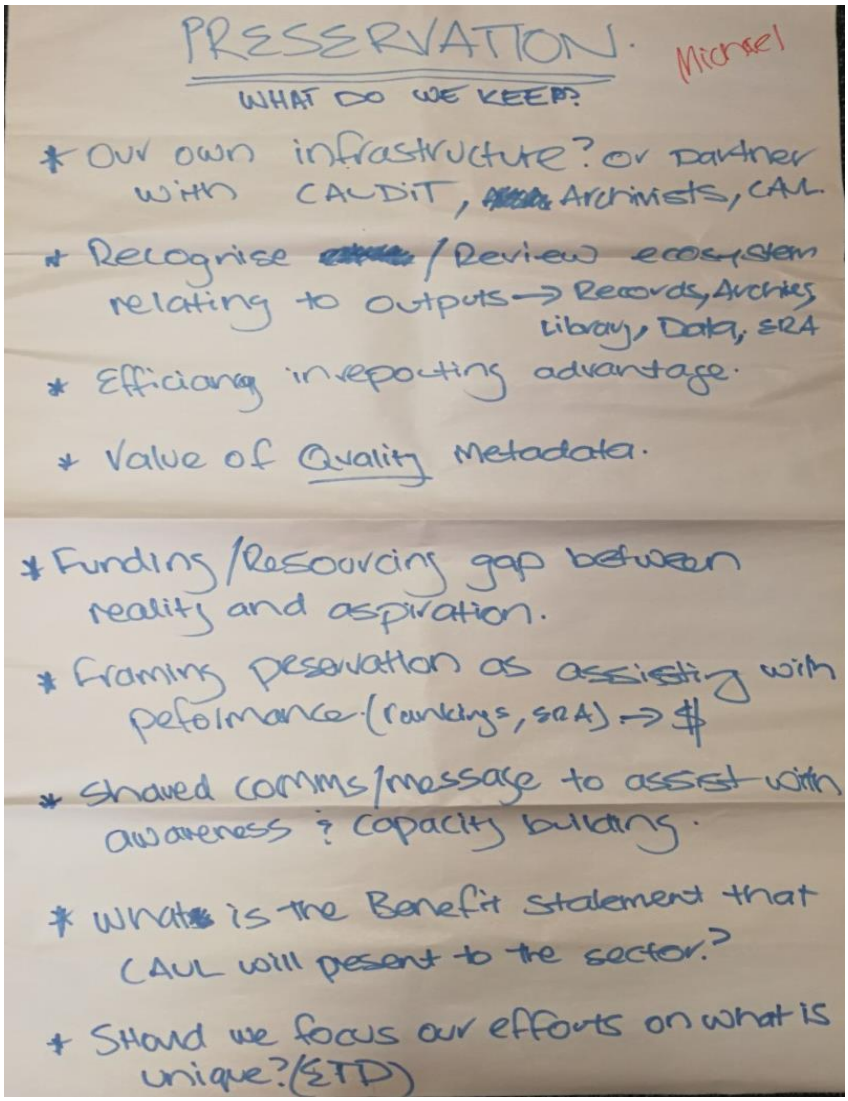
General support for all recommendations – make sense

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Preservation

What Do We Keep?

-by Michael's group



- Our own infrastructure? Or partner with CAUDIT, Archivists, CAUL
- Recognise/Review ecosystem relating to outputs -> Records, Archives, Library, Data; ERA
- Efficiency in reporting advantage
- Value of Quality metadata
- Funding/Resourcing gap between reality and aspiration
- Framing preservation as assisting with performance (rankings, ERA) -> $

- Shared comments/message to assist with awareness? capacity building
- What is the benefit statement that CAUL will present to the sector?
- Should we focus our efforts on what is unique? (ETD)

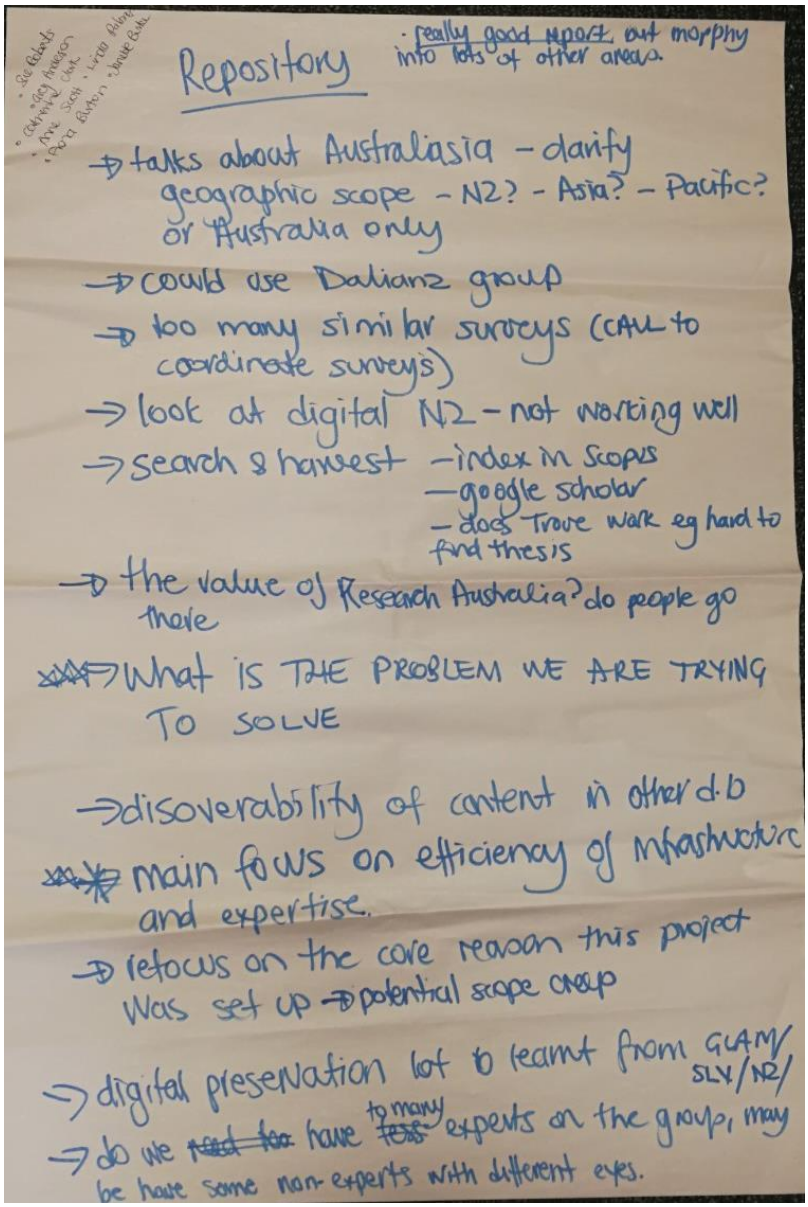
CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Repository

Really good report, out mopping into lots of other areas

By Sue Roberts, Greg Anderson, Catherine Clark,

Anne Scott, Linda Palmer, Fiona Burton, Janette Burke



- Talks about Australasia – clarify

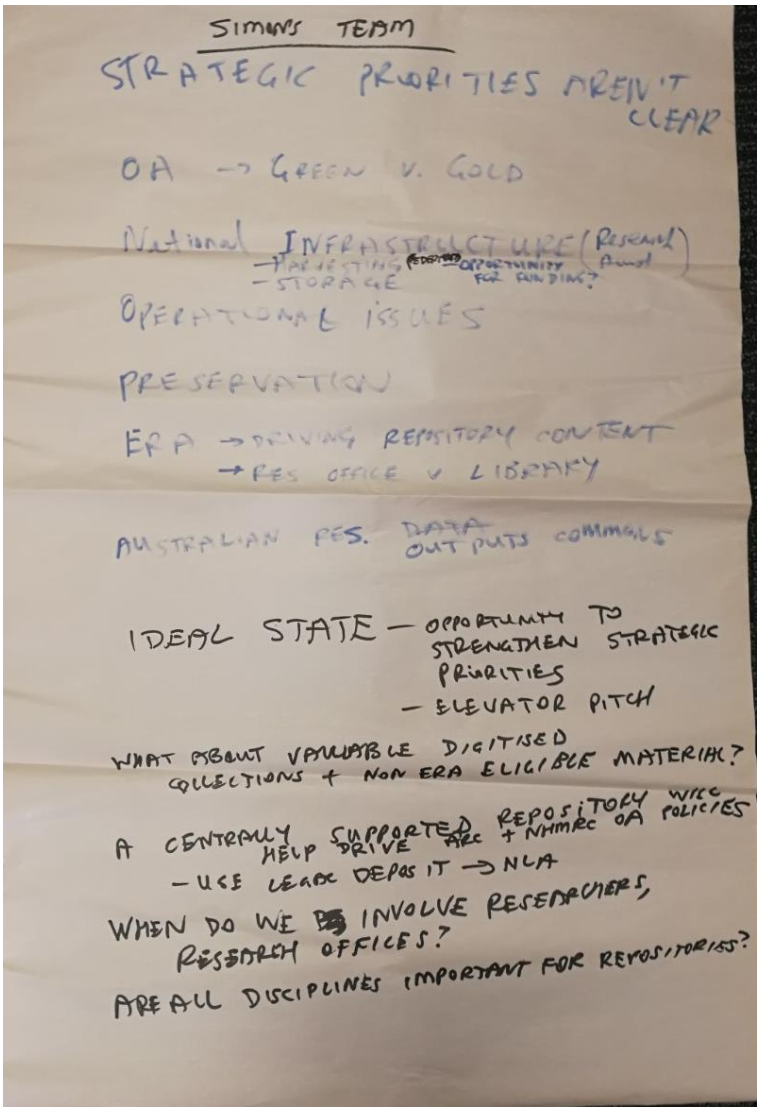
  Geographic scope -NZ? –Asia? –Pacific? Or Australia only
- Could use Dalianz group
- Too many similar surveys (CAUL to coordinate surveys)
- Look at digital NZ - not working well
- Search & harvest - index in scopes

- Google scholar
- does trove work eg. Hard to find thesis

- The value of Research Australia? Do people go there?
- What is the problem we are trying to solve
- Discoverability of content in other db

  * Main focus on efficiency of infrastructure and expertise
- Refocus on the core reason this project was set up -> potential scope creep
- Digital preservation learnt lot to from GLAM/SLY/NZ
- Do we have too many experts on the group, may be have some non-experts with different eyes

Simon's Team



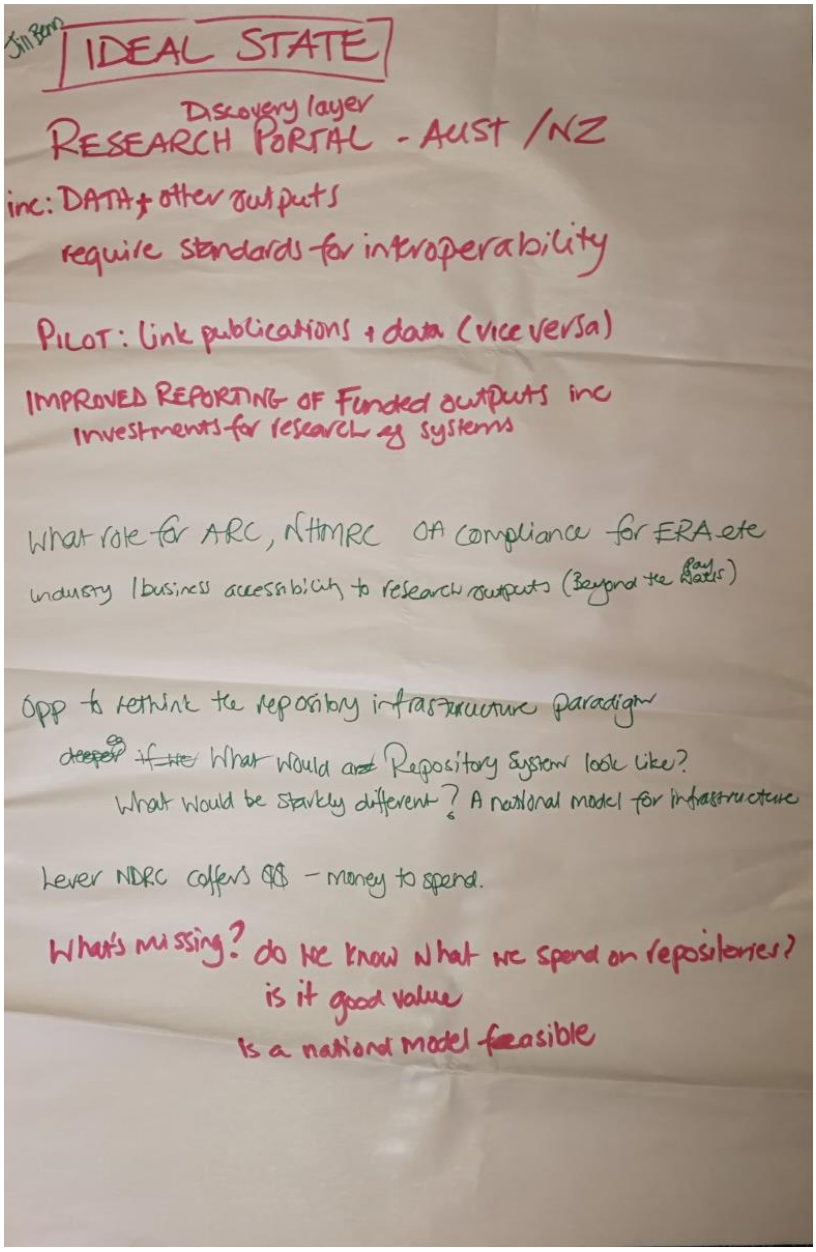
- Strategic Priorities Arnt Clear
- OA -> Green vs. Gold
- National Infrastructure (Research Australia)
  - -Harvesting - Opportunity for funding?
  - -Storage
- Operational Issues
- Preservation
- ERA -> Driving repository content
  - -> Res office vs. library
- Australian research outputs commons
- Ideal State - opportunity to strengthen strategic priorities
  - - Elevator pitch

- What about valuable digitised collections + Non EAR eligible material?
- A centrally supported repository will help drive are + NHMRC OA policies -- use lease deposit it -> NLA
- When do we involve researchers, research offices?
- Are all disciplines important for repositories?

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

IDEAL STATE

By Jill Benn's Group



Discovery layer

Research Portal –AUST/NZ

Inc: Data + other outputs

Require standards for interoperability

Pilot: link publications + data (vice versa)

Improved reporting of funded outputs include investments for research eg. systems

What role for ARC, NHMRC OA compliance for ERA etc

Industry/ business accessibility to research outputs (beyond the paywalls)

Opportunity to rethink the repository infrastructure paradigm

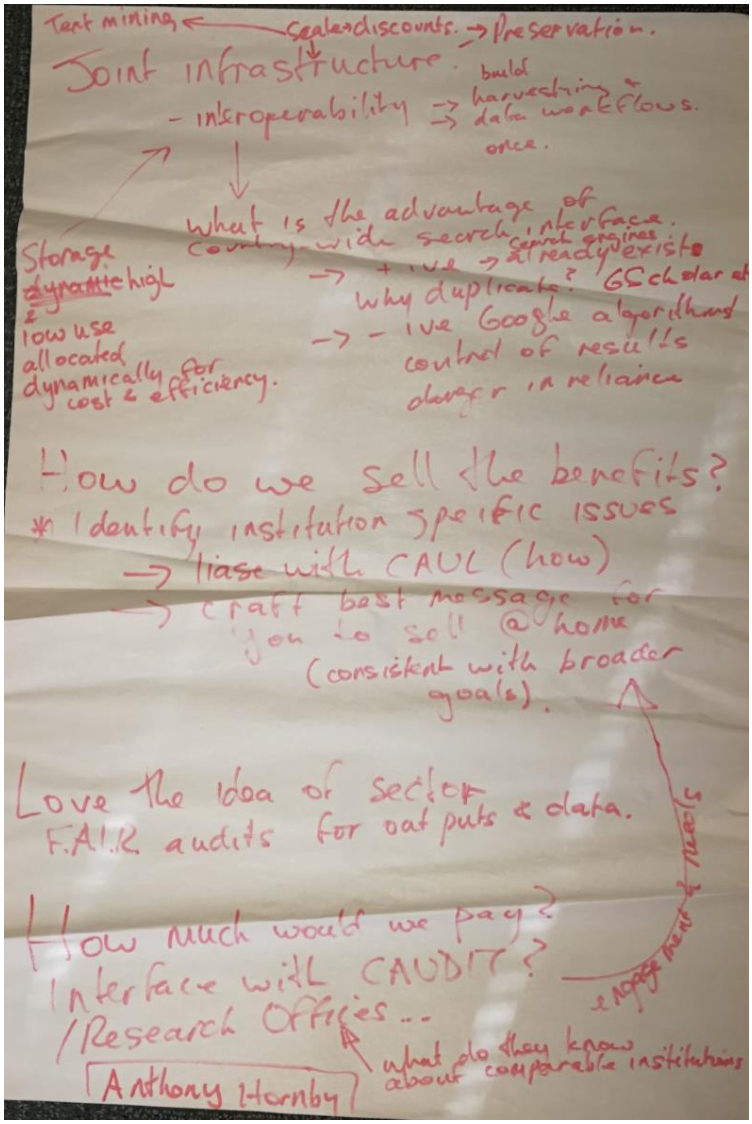
What would be starkly different? A national model for infrastructure

Lever NDRC coffers $$ - money to spend

What's missing? Do we know what we spend on repositories? Is it good value? Is a national model feasible?

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

By Anthony Hornby's group



Text mining ← Sale → Discounts → Preservation

↓ ↗

Joint Infrastructure

Inoperability → harvesting

→data Workflows once

Storage high & low use allocated dynamically for cost & efficiency

What is the advantage of countrywide search interface?

→ +If Already exists, why duplicate

→+If google algorithms control results a danger in reliance?

How do we sell the benefits?

*Identify institution specific issues

→ liaise with CAUL (how)

→Craft best message for you to sell at home (consistent with broader goals)

Love the idea of sector
F.A.I.R audits for outputs &data
How much would we pay? Interface with CAUDIT/Research Offices… → what do they know about comparable institutions

Feedback from Harry's Group

Size of the document – too large, the group thought that it needed a proper summary with themes.

Trove as starting point for research Australia? The group didn't think that this was a good place to start.

[Recommendation 11] Some institutions can have guidelines but not a policy, a template is fine.

The future of CAIRSS was seen as important, the group supported all the recommendations in principle.

[Recommendation 18] is this about page 44, what is it actually asking people to endorse?

How to provide feedback after the CAUL meeting? The group asked for more opportunities to give detailed feedback.

How will the work be taken forward after the report is finalised, and who will do this?

The report mentions lots of statistics for repositories but are these based on an international standard like IRIS?

The group thought that recommendations could be collapsed and condensed.

The group thought that aiming for everything to have a CC attribution might not be a realistic goal.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au

Appendix 9: Feedback provided in 2019

Review of Project Products

At the CAUL Council 2018/2 (September) meeting, attendees were asked to provide feedback to the draft project report using butcher paper notes. Feedback is provided as Appendix 8.

A secondary feedback process was undertaken in February – March 2019 via a Zoom videoconference and online survey. Feedback from this process is provided here as Appendix 9.

CAUL (Council of Australian University Librarians)
+61 2 6125 2990 | caul@caul.edu.au | www.caul.edu.au