Abstract: Collection, curation and reporting of research outputs is a Government-driven requirement for all Australian universities. While most universities have Institutional Repositories (IR) in place, only a handful have developed a fully integrated model where all research outputs are submitted to the repository and all necessary assessment data is captured through the deposit process. This paper describes the development of the CSU institutional repository - Charles Sturt University Research Output (CRO) which succeeded in collecting all reportable research publications for 2007 within its first year of operation. The successful development of CRO was dependent on the ability to create a flexible system which could capture and manage known data requirements and accommodate the addition of others. Asking and achieving more from the 'out of the box' repository software relied on effective partnerships not only between Charles Sturt University and the UNILINC library network, but also with the support of the system vendor and importantly the local research community. The achievement of such a model - that is sustainable, scalable, and integrated within the institutions community and research processes - resulted from effective leadership, collaboration, innovation and partnerships across the institution and beyond.

Author(s) Address: sbishop@csu.edu.au  kasmith@csu.edu.au

URL: http://www.information-online.com.au/


http://researchoutput.csu.edu.au/R/?func=dbin-jump-full&amp;object_id=10562&amp;local_base=GEN01-CSU01

CRO identification number: 10562
Abstract

Collection, curation and reporting of research outputs is a Government-driven requirement for all Australian universities. While most universities have Institutional Repositories (IR) in place, only a handful have developed a fully integrated model where all research outputs are submitted to the repository and all necessary assessment data is captured through the deposit process.

This paper describes the development of the CSU institutional repository - Charles Sturt University Research Output (CRO) which succeeded in collecting all reportable research publications for 2007 within its first year of operation. The successful development of CRO was dependent on the ability to create a flexible system which could capture and manage known data requirements and accommodate the addition of others. Asking and achieving more from the “out of the box” repository software relied on effective partnerships not only between Charles Sturt University and the UNILINC library network, but also with the support of the system vendor and importantly the local research community.

The achievement of such a model - that is sustainable, scalable, and integrated within the institution’s community and research processes – resulted from effective leadership, collaboration, innovation and partnerships across the institution and beyond.

Introduction

Australian Universities have developed or are developing open access institutional repositories to showcase research outputs of the institution and to meet government requirements that publicly funded research becomes accessible to all. Open access is gaining momentum in the eResearch environment, with initiatives underway to extend access to research data in addition to research publications. The Federal Government’s Research Quality Framework (RQF) required all universities
to have an institutional repository by the time the RQF assessment exercise was due to commence in April 2008. The RQF was replaced by ERA (Excellence in Research for Australia) in 2008, however the imperative to make research outputs openly available for use and assessment is becoming an important component of the scholarly communication process.

The RQF was the principle driver for many universities to fast-track development of their institutional repositories. This was certainly the case for CSU. The University had been investigating options for creating an open access repository for some time but had not taken steps towards implementing one. This paper outlines how a tight deadline, Government reporting requirements and limited internal resources, a situation which was initially daunting, created an opportunity for CSU to build an integrated repository and research management system. The success of the CSU model is predicated on a University-wide mandate and collaboration: within the University; with our hosted service provider, UNILINC, which also became the development partner; and with the product software supplier, ExLibris.

Background

In mid 2006 Charles Sturt University’s (CSU) Senior Executive Committee assigned responsibility to the Division of Library Services to lead a project to develop and implement an institutional repository, to be operational by early 2007. It was known that the RQF would commence in 2008 and, to meet its requirements, CSU would need an open access institutional repository in place. Like all other Australian universities at that time, CSU knew nothing about the technical requirements of the RQF. The project team worked on the assumption that whatever was developed must conform to international standards for data capture, storage, retrieval and output.

A tight delivery deadline and a lack of internal resources to develop and maintain an onsite repository within the assigned timeframe drove the initial decision to seek an externally hosted repository product. CSU is not unique in having chosen a hosted repository service rather than adapting open source software. Many of the reasons for CSU taking this path are well articulated by Sutherland & Hopkins (2007) when discussing Bond University’s experience. Essentially the resource intensiveness of other open source applications in development at CSU at the time left no room for adapting another open source product. By outsourcing the technical support and storage, the project team could concentrate on defining the University’s requirements and a model for achieving and populating an integrated repository and research management system. The repository would address the RQF and, more immediately, the University’s research reporting requirements under the Higher Education Research Data Collection (HERDC) process.

The brief for the repository project team was to develop a system to capture all of CSU’s compliance-driven reporting, be flexible enough to address RQF specifications (then unknown) and provide management information on CSU’s research publications. The system would also capture, store and make available, through open access harvesting, the published research outputs of CSU to promote the University’s research strengths. At the same time the University was seeking a system which could expand into other areas of digital collection such as eTheses and research data sets. ePress capabilities were included in the brief, to be explored in a later phase of the project. After an
extensive assessment process CSU selected ExLibris’s DigiTool software, to be hosted under licence by UNILINC Limited. UNILINC provides CSU’s shared library system and other online services with both organisations having a long history of working together.

Uncertainties about the specific requirements of the RQF delayed the project and it was not until August 2007 that implementation of the software commenced in earnest, considerably later than the initial project delivery deadline. At this point UNILINC staff members joined the project team meetings and joint development of CRO began.

When the project began, no other Australian university had implemented a DigiTool institutional repository although several were in the process of developing one for RQF purposes (University of Melbourne, Curtin University of Technology, University of Western Australia). In Australia and overseas there were numerous examples of sites which had implemented DigiTool collections. However none (that could be discovered) had been developed to the degree proposed by this project. CSU required an open access, self-archiving, research repository to capture research management metadata in addition to descriptive metadata and objects. The initial thought was to implement an “out of the box” version of DigiTool repository software to store CSU research objects with the intention of gradually modifying it in the second year of operation. It became apparent that considerable modification would be required from the start if the repository’s development was to be tied to HERDC and address the draft RQF Specifications, released in July 2007. Delivery of a “live” repository was required as soon as possible and before the end of 2007.

The CSU Model

Support for the repository by the University’s Senior Executive has ensured it has been given high priority within CSU. The Library was selected to lead the project and has strong support and input from the Deputy Vice Chancellor Research who is a member of the repository steering committee. CSU’s research office, the Centre for Research and Graduate Training (CRGT), is a significant stakeholder and client. The Library and the CRGT share administrative roles in the repository according to each section’s expertise, following jointly developed procedures. This arrangement has been key to the success of the CSU model. Ultimately the Library is responsible for the overall management and administration of the repository, however, this is dependent at all times on the collaborative relationship established with the CRGT and the researchers with whom the CRO administrators interact.

<table>
<thead>
<tr>
<th>CRGT</th>
<th>Library</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check submission data completeness</td>
<td>Confirm citation details for published item</td>
</tr>
<tr>
<td>Check publications meet CRO inclusion criteria</td>
<td>Verify and apply open access rights and permissions</td>
</tr>
<tr>
<td>Verify eligibility of HERDC and other compliance – driven</td>
<td>Negotiate open access permissions</td>
</tr>
</tbody>
</table>
Table 1: CRO Shared Administrative Responsibilities

<table>
<thead>
<tr>
<th>Submissions</th>
<th>With Publishers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>File formatting, application of external links etc</td>
</tr>
</tbody>
</table>

The release of the *Institutional Repository Checklist for Serving Institutional Management*, Carr et al (2008) has provided reinforcement for CSU’s model. The checklist, published in April 2008, was developed as a response to the United Kingdom’s Research Assessment Exercise (RAE), a precursor to, and model for, the RQF. The CSU project team decided to apply it as a ‘health check’ and tested our institutional repository, CRO (CSU Research Output), against the 13 criteria. CRO is 100% compliant with several criteria, notably items 10-13, and meets most of the others to a substantial degree. Results are contained in Appendix 1.

The *ARROW HERDC Working Group Interim Report* (May 2008) makes interesting reading in light of the direction CSU chose to take in 2006 by developing CRO for HERDC submission. The ARROW report proposes potential models for integrating research management data and institutional repositories. It underscores many of the challenges CSU has faced in developing CRO to capture and report HERDC data. The project team has encountered many of the issues raised, including defining shared workflows across administrative units, data review and verification, and duplication of entries to name some. By tying the development of CRO to HERDC, a process already well entrenched and understood at CSU, the impact of introducing a new process to researchers has been minimised. Submission of HERDC data is mandatory and the CRGT provides an additional financial incentive to researchers who submit their publications data. By adding the requirement to submit a digital version of the research paper as part of the HERDC process the repository achieved a high percentage of full text content in its first six months of operation, over 1,000 submissions by 30 May 2008.

Much has been written on the difficulty of getting content into institutional repositories when deposit is voluntary. Considerable effort is required to promote the repository and encourage researchers to submit their work, even then success may not follow. It is relatively easy to transfer research publications metadata from research office systems to the repository but more difficult to recruit pre-print versions of papers acceptable for open access inclusion. Even the best promoted repositories, with a high level of academic buy-in struggle to achieve full text content if the process is voluntary. Davis & Connolly (2007) discuss the imbalance of content in repository collections at Cornell University where there is no overarching mandate for deposit. While some faculties at Cornell do mandate deposit and these collections show steady growth, the collections of others remain empty or poorly populated. Ferreira et al (2008) point to the significant jump in content, around 71% increase in deposits, at the University of Minho when mandated submission coupled with financial incentive was introduced. In Australia, the University of Tasmania’s Professor Arthur Sale has consistently stated the need for mandatory deposit and has presented compelling evidence.
to illustrate that it is the most successful strategy for capturing content. From a researcher’s point of view there are many understandable barriers to voluntary self-archiving. Researchers are time-poor, often juggling the dual pressure of research and teaching commitments, they may not retain their own copies of pre-prints after submission to publishers, some are concerned that the pre-print version will be perceived as being of lesser quality than the published version. They may be worried about the intellectual property implications of depositing to an open access repository and be uncertain about what rights they have signed over to publishers. Many are yet to be convinced of the benefits of archiving their research in an institutional repository. Changing these perceptions takes time and effort, often the most effective tool for influencing researchers is the word-of-mouth evidence of colleagues who have experienced increased interest in and/or increased citation of their research as a result of exposure in an open access repository.

CSU’s mandatory submission policy states:

*All CSU staff are required to submit their peer-reviewed pre-publication manuscripts for the DEST Higher Education Research Data Collection (HERDC) as an electronic file into CRO from 2007 onwards. Items submitted will constitute the CSU open access repository and all items are retrievable through open access search engines. The CRO Repository is searchable through the CSU web site and full text items are available for download. CRO will be used to determine funding under the Publications Grants Scheme and to provide data to the Promotions Committee. For 2007, citation details will be used to calculate DEST points (in the absence of the appropriate pre-published version of the publication) however everything possible should be done to place a pre-print version on CRO. For 2008 publications, the pre-print draft must be entered onto CRO to count for Publication Points.*


By mandating document submission, tied to the HERDC process, CSU is able to capture a much higher proportion of full text, suitable for open access archiving while it is still “fresh”. This process is further reinforced by the CRGT’s payment of research publication grants for HERDC submissions. In its first 6 months of operation CRO has grown to having over 1,000 research records, some 70% of these have full text open access compliant documents attached (Figure 3) and 74% of those are able to be viewed (Figure 4). Over 50% of total content in CRO is “unlocked” for open access viewing.
While CSU is not a research intensive university, the University’s reportable research output has increased steadily since 2001. The graph (Figure 5) indicates publications growth in the period 2001-2007. For the repository to grow sustainably, researchers need to be able to self-archive directly and as soon as possible after their work is accepted for publication (see Sale (2006) *Acquisition of open access research articles*). The CRO submission forms were designed to be as intuitive as possible, minimising the need for extensive training and support. The relatively few requests for assistance or
adverse comments during the 2007 HERDC process support this decision. University funding models changed for 2008 HERDC, with incentive payments to researchers who submit to CRO, made available soon after submission. This should encourage more timely deposit and help to mitigate the last minute rush for entering papers prior to the annual HERDC deadline.

Figure 5: Increase in the number of publications submitted for HERDC between 2001 and 2007

**Approach**

Selection of software, implementation and development of CRO has been run under applied project methodology following the PMBOK (Project Management Body of Knowledge) adapted for CSU by the University’s Project Service Centre. The project team represents key stakeholder groups: Library, CRGT, academic staff, Information Technology. By following project methodology the project team was able to identify other initiatives, planned or underway at CSU, which could have been potential dependencies for this project, ensuring that there was no duplication of effort or conflict in allocation of resources. The project approach provided a sound structure for scoping and defining the University’s business and functional requirements for an institutional repository, and established ranked criteria from which the software selection was made. An important criterion for the selected software was that it must support international standards to enable data migration to a different repository system hosted by CSU at a later date (should it be required) and to enable data sharing and transfer within CSU and to external systems such as the proposed RQF Information Management System (IMS), either by system to system protocols or through middleware applications.
To achieve this model, the project was attempting to integrate two administrative processes: HERDC data collection and the collation and bibliographic description of research papers. While at first appearing closely related, as work processes evolved, the differences in approach to these tasks by the research office and the Library became more obvious. Despite the challenges, the benefits of merging these processes outweighed the problems, with the main advantage being guaranteed content for the repository using a process already part of the lives of academics.

In the spirit of sharing scholarly communication through open access, the institutional repository provides bibliographic information about research outputs linked, in most cases, to a full text digital version of the work. Administrative tasks associated with this process include: collation of documents; confirmation of accurate bibliographic information and other descriptive metadata; badging the document so it is identifiable as CSU research regardless of its mode of discovery, ie via open access search engines or as an embedded link in other documentation; identifying whether the copyright holder will allow open access and under what conditions. Technical requirements include: secure and stable storage of the digital objects and their associated metadata; application of metadata standards to create a data structure that can ensure reliable and repeatable open access retrieval of the papers.

HERDC is the mechanism for reporting data about research publications in the Australian higher education sector. In addition to the descriptive bibliographic metadata normally held in the repository, administrative information about the author and the work is required for HERDC reporting and for CSU’s internal management processes. Data collected by CRO for this purpose includes: staff ID number; phone number; faculty, school, research centre affiliation; coded information about the research discipline area e.g. Australian and New Zealand Standard Research Classification (ANZSRC) Field of Research (FOR) codes. Elements of the information collected for HERDC are confidential and remain hidden in the public interface of CRO but are able to be retrieved through reports run by the administrators. Collation of this data and other tasks which differ from those in non-integrated repositories include: verification of the article for HERDC compliance, affirmation of correct bibliographic citation details and the author’s affiliation with the University; allocation and recording of “publication points” according to Government criteria and the allocation of funding within the University based on the information collected. This process benefits the repository by quality assuring the accuracy of the publications data submitted. Integrating the additional data requirements into the standard repository software and enabling the creation of meaningful reports for HERDC has been the most problematic area in this combination of processes.

**Implementation**

At the outset various aspects of the software had to be customised to create:

- Streamlined deposit process for academics
- Authentication for the submission process
- Efficient workflows to approve and quality assure items
- Data structure for reporting and open access harvesting
To meet the project deadline, the repository was implemented very quickly with the software loaded to a UNILINC server at the end of August 2007. Training was delivered by the vendor to UNILINC staff and the Library’s repository administrator in September. In the three months between September and the end of November a CRO Metadata sub-committee was convened and the necessary Dublin Core fields identified. The submission forms were created and tested with a small group of researchers who provided feedback and suggestions, and authentication was set up. CRO opened to CSU researchers for their 2007 HERDC submissions in mid-December. To have achieved a “live” repository in such a short time from start up to first deposit is a reflection of the commitment by all involved in the project team and by CSU’s external partners, UNILINC and through it, ExLibris. It was a vindication of running the project under established project methodology to keep it on track and ensure there was no scope creep to distract participants or introduce unnecessary delays.

The administrative workflows developed further as the first submissions were received. Submissions were gradual to begin with, corresponding with the December-January holiday period, which gave the administrators time to assess the order in which processing would be done. Closer to the submission deadline (set for CSU researchers at 30 May) the trickle of submissions became a torrent. The ensuing work load required a rethink of the workflows to streamline processes such as verification of papers against the HERDC and CRO inclusion criteria. The post-implementation review, conducted after the 2007 HERDC submission, alerted the project team to the challenges and workload issues associated with integrating the CRGT and Library processes. The review indicated improvements to be made to the forms, order of workflow tasks and FAQ type information required for the 2008 HERDC process.

Quick implementation is not without its drawbacks. The project team and UNILINC learned much by running the integrated repository in an operational environment to capture and report the University’s 2007 HERDC data. The biggest challenge was to extract data from the repository software as it has not been designed to deliver reports of the complexity CSU required. UNILINC technical staff were able to create SQL reports to extract the data. However it quickly became apparent that the Dublin Core data structure and the local fields created to accommodate HERDC data, inhibited the reporting process.

**Authentication**

Author self archiving is something CSU took for granted as researchers have completed online forms for HERDC submission for some time. Since CRO is hosted off campus and outside the CSU information technology environment, it was critical that the submission forms and authentication worked with minimum disruption to the process the researchers were familiar with. It has been enlightening to discover that authentication is less of an issue for some organisations. For example Bond University (Sutherland & Hopkins, 2007), McGill University in Canada (another DigiTool repository) and a number of other universities do not require academics to complete their own submission to the repository. Submission by the repository staff was not an option at CSU where resources are not available.
Discussion with McGill University about their work processes revealed that submitter authentication was not going to be as straightforward as first thought. It was anticipated that a single sign-on remote authentication system would be used, based on each CSU staff member’s University sign-on. After some considerable frustration at many failed attempts, UNILINC staff were able to demonstrate there was a bug in the DigiTool Deposit module preventing this approach from working. Despite there being numerous installations of the software around the world such a significant problem had not been found earlier, presumably because few universities had instituted direct deposit by academics. The combination of CSU pushing its requirements, UNILINC seeking technical solutions and ExLibris responding to the feedback ensured that the problem was highlighted and solved for the rest of the DigiTool community in a subsequent service pack. To meet project deadline CSU opted for a workaround authentication solution, using the separate Library system authentication process which was familiar to CSU users and compatible with the repository software.

Submission

A simple, streamlined, well designed process for submitting information is seen as essential to ensure that clients submit their information into the system. The plain vanilla DigiTool submission form was customised to capture the information required for HERDC. This made the form quite long and detailed. The forms originally allowed for entering up to three authors with a separate, single field for all additional authors. CSU needed to expand the author fields to allow for up to eight authors because so many papers are collaborative (for the relatively few papers with more than eight authors, the CRO administrators enter the additional data directly into the administration module). Large numbers of authors associated with a single paper created difficulties for reporting if they were not entered in individual fields. Fields need to be available for capturing all of the authors even those not affiliated with CSU. The additional fields make the form look unwieldy which can be off putting to users. However the trade-off is that the information only needs to be filled out once. Thus enabling researchers to report their HERDC activities and submit to the repository in a single process. At time of writing, many universities with separate research systems and repositories have two separate submission processes, requiring duplication of data entry and associated potential for data inconsistency in the separate systems.

Several researchers requested that the forms automatically load their personal information with data taken from the University’s personnel system. This is currently not possible, but improvements made to CRO’s data structure have the potential to enable data sharing between the repository and the University’s administrative systems at a later date.

Populating CRO

Collecting of the correct version of the paper is another potential issue raised by the ARROW HERDC Working Group. To ensure deposit of open access compliant content in the repository the CRO submission policy asks for the final draft (as accepted for publication) to be submitted. To prevent confusion and difficulty with the new system, the published version was not required. This has led to difficulties in the administration process as the published version needed to be followed up.
separately, to verify citation details and compliance with HERDC criteria. A potential solution is to require a DOI or URL where available for the 2008 collection so the published version can be easily located.

**Data structure**

It became apparent when the data was processed for reporting to HERDC that CRO’s Dublin Core data structure was not flexible enough to create accurate reports direct from the system. The DigiTool submission forms map information to Dublin Core fields and the CRO Metadata subcommittee needed to create local (unique to CRO) fields to collect the HERDC-specific data.

Dublin Core did not meet CSU’s needs when it came to producing reports for HERDC because, without structured subfields, it was impossible to accurately link individual authors to information such as their email address, research discipline codes, faculty or research centre. This was a significant problem as this data is crucial for reporting and assigning research funding to schools and research centres. The solution for 2007 HERDC was to dump data to an external spreadsheet for manipulation into HERDC format. The long-term solution has been to implement MARC21 as CRO’s internal data structure. In July 2008 the CRO Metadata subcommittee re-convened to review the Dublin Core fields, and map them to appropriate MARC 21 fields and subfields. The mapping was based on the Library of Congress Crosswalk standard ([http://www.loc.gov/marc/marc2dc.html](http://www.loc.gov/marc/marc2dc.html)), with additional mapping of locally defined CRO Dublin Core fields to MARC 21 fields and subfields. The submission forms still use Dublin Core fields which are cross-walked into MARC 21 and are able to be translated back to Dublin Core, if required, for export. Figure 4 is a snapshot of CSU Dublin Core local fields with their translation to MARC 21 showing MARC subfields.

<table>
<thead>
<tr>
<th>Dublin Core Fields</th>
<th>MARC 21 field / subfield</th>
<th>Journal Article Form Labels (as at 9/7/08)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc:creator@dcterms:csuResearchCentre</td>
<td>700 1#$m</td>
<td>Research centre</td>
</tr>
<tr>
<td>dc:creator@dcterms:csuDestPoints</td>
<td>700 1#$n</td>
<td>Publication points claimed by this author</td>
</tr>
<tr>
<td>dc:creator@dcterms:csuEmail</td>
<td>700 1#$f</td>
<td>Email</td>
</tr>
<tr>
<td>dc:creator@dcterms:csuFaculty</td>
<td>700 1#$k</td>
<td>Faculty</td>
</tr>
<tr>
<td>dc:creator@dcterms:csuGender</td>
<td>700 1#$g</td>
<td>Gender</td>
</tr>
</tbody>
</table>

Table 4: Sample of Dublin Core local field conversion to MARC21 within CRO.
Collaborative Relationships and beyond

UNILINC and ExLibris had originally explored the idea of providing a hosted “out-of-the-box” repository service based on the DigiTool software platform as early as 2006. However, this was not formalised and did not go anywhere at the time. The CRO project was the impetus to revisit the original hosted service concept. It was recognised by all three parties that this was going to be a complex project to deliver. Adapting the “out of the box” version of DigiTool to meet CSU requirements necessitated extensive liaison with the software vendor, ExLibris, to advise on implementation and assist with the changes. UNILINC provides a range of library services, many of which are based on ExLibris software. Leveraging UNILINC’s existing relationship and established contacts with ExLibris has enabled the project to meet most of its objectives to date. UNILINC provides pivotal liaison between the CSU project team and ExLibris. UNILINC staff translate CSU’s requirements into technical specifications and work with DigiTool developers to assess feasibility and/or discover alternative workarounds to issues which have emerged with the development of CRO. This is well illustrated in the work done to resolve DigiTool’s reporting shortcomings by realigning the data structure from Dublin Core to MARC 21. Since the project began, the tripartite collaborative arrangement has been formalised by service level agreements, between CSU and UNILINC; and by a licensing agreement between UNILINC and ExLibris.

The ARROW HERDC Interim Report confirms that any integrated repository software solution, whether open source, proprietary, or hybrid, requires significant customisation. There are a number of issues to be worked through to achieve a fully integrated research repository such as that aimed for by the CRO project. Key to the project’s success have been the clear objectives of the project brief and commitment by the three partners to work together to achieve these. ExLibris continues to refine and develop DigiTool software. It is heartening to learn that some of the developments will address issues the CRO development team has raised in the course of this project. Proposed developments include; improved Open Archive Initiative harvesting, enhanced reports and Resource Discovery services, as well as preservation capabilities. ExLibris reports in the July Ex Librarian Newsletter (2008) that the company is developing a new open-platform program to enable the product users and vendor to work together to extend and refine the product in a more open and flexible framework to better meet local requirements. This will suit the Australian user community with its history of effective collaboration.

That Australians are early adopters of technology is a characteristic supported by this project’s discovery that what CSU was trying to achieve with CRO did not appear to have been considered elsewhere by other DigiTool users. By pushing the envelope, in collaboration with the software developer, the capacity of DigiTool repository software is likely to be further enhanced, not just to CSU’s benefit but to the benefit of other users and prospective clients. In addition to the CRO integrated research repository there are considerable innovative developments within the Australian DigiTool community covering cultural heritage preservation collections, parliamentary library collections, eTheses and course materials.
Community working together

An unexpected, positive and mutually beneficial outcome of the CRO project was the advent of the DigiTool Working Group (DTLWG). This informal group was brought together around the same time as the first installation of DigiTool at UNILINC. It seemed a good idea to make contact with all the then DigiTool sites in Australia particularly with the RQF in mind. This proved very successful, even if just for mutual support during the quite hectic times around the RQF. With the demise of the RQF it was not expected that the DTLWG would necessarily continue. In reality, this informal group has grown considerably and evolved into an effective community for sharing ideas, experiences and expertise. This spirit of collaboration is best exemplified by the willingness of members to show in detail their individual implementations to the rest of the group. The DTLWG is not affiliated to the system vendor, nor does it belong to the formal vendor user group community. Its existence is more akin to the communities that grow up around open source development.

Conclusion

The project has shown that integrating the institutional repository into the academic and business processes of the University can achieve both collection of data for government reporting and a content rich repository. The ground-breaking collaboration between the CRGT and the Library has meant the new process has been championed by the Deputy Vice Chancellor Research, a significant stakeholder, whose influence on the CSU researcher community and carriage of a University-wide mandatory submission policy has ensured the growth of repository collections. The Library has led the project to unite the vision of an integrated repository with the practicalities of implementation. It has contributed the information management skills of its staff to find ways to stretch the institutional repository beyond its traditional boundaries. UNILINC has brought its technical expertise and long history of serving the library community to provide an innovative, MARC 21 based solution to resolving reporting functionality, not available in the original repository software. UNILINC’s long-standing relationship with the repository system vendor and ExLibris’s responsiveness to software issues illustrates that timely software development is not necessarily the sole preserve of the open source community.

CRO has become an integral part of the University’s systems architecture. The repository is the authoritative data source for CSU research publications information. It has the potential to share this data with other CSU systems, to be reused for different purposes.

Building an integrated institutional repository is very dependent on the existing structure and systems in place at each institution. The ARROW HERDC Interim Report offers four potential models for achieving an integrated repository, taking into account differing institutional structures. CSU has been well placed to implement Model 4, a fully integrated repository. This is due to support from the University Senior Executive, willing collaboration between the two administrative areas, Library Services and the Centre for Research and Graduate Training. An additional advantage for the project team was that it was not constrained by an existing proprietary research management system, the previous system being an in-house database which needed replacing. Measuring CRO against the
Institutional Repository Checklist (Carr, et al 2008, see Appendix 1) confirms this aim has been largely achieved. The CSU model will not work for everyone however it is hoped there are elements in this approach which can be used or adapted by other repository developers.
References


Appendix 1

CRO performance against the Institutional Repository Checklist for Serving Institutional Management.

Criteria 1: The repository must uniquely identify each local (institutional) author for each item.

CSU does not maintain an author authority file however CRO collects Staff ID number as a consistent tag for each author.

Criteria 2: The repository must contain a complete, accurate and updateable list of all the academic staff, faculty, researchers and professors in the institution.

At this stage CRO stores information about research-active staff, required for HERDC reporting i.e. those who are currently publishing. At a later stage this data may be able to be shared with other administrative systems in the university, or ingested from those.

Criteria 3: The repository must track the affiliation between individuals and their departments or research groups.

The MARC21 data structure in the repository allows collection and retrieval of author affiliation information.

Criteria 4: The repository must be able to handle content-free items

Records without attached content are included in CRO

Criteria 5: The repository must be able to manage dark items

CRO archives metadata which is not displayed or searchable in the public interface, papers and collections are or can be locked from public view if they do not comply with open access rights or for any other reason.

Criteria 6: Reports generated by non-system staff

This is not possible at this stage. It is planned to set up and distribute regular reports on CRO submissions for research administrators, heads of research centres and faculty. Ad hoc reports will be created as required.

Criteria 7: The repository must provide quality assurance processes that can handle high throughput

Quality assurance is undertaken as part of the approval process, workflow procedures and checklist boxes in the administrator’s module ensure QA verification is completed prior to making content live.
Criteria 8: The repository must be able to support high-throughput deposit processes

The deposit module allows for high throughput. Rate of throughput is dependent on rate of self-archiving by researchers. All items are verified by administrators to ensure they are eligible to earn publication points. Verification may take longer in peak periods however this does not impede the deposit process.

Criteria 9: The repository should have a policy for dealing with ‘very low quality’ items deposited by schools or individuals.

The CRO Submission Policy and associated inclusion criteria specify allowable content.

Criteria 10: The repository must support metadata fields to tie its records in with other information systems

The software is compliant with international metadata standards and is able to output data in a variety of formats including Dublin Core, MARC 21, XML, CSV

Criteria 11: The repository should involve (and should be involved by) senior management and administrative committees to guarantee institutional embedding.

CRO gets five stars for this. Supported as a high priority University project by the CSU Senior Executive, strong stakeholder support from DVC Research.

Criteria 12: The repository needs a strategy for maximising ‘full text deposit’ along with metadata records.

CRO gets five stars for this as evidenced by 2007 HERDC and future mandatory reporting. The legacy data (2001-2007 HERDC metadata records) contains metadata only with a University wide promotion strategy to obtain the papers.

Criteria 13: The repository needs to support the practices and assimilate the legacy systems of fiercely independent departments and schools.

Tying CRO to centralised HERDC reporting and the lack of proprietary research office system has meant that this has not been a problem at CSU.