"Too many systems, too little time": integrating an eprint repository into a University publications system

Julie Woodland
Senior Librarian
Library and Information Service
Curtin University of Technology
j.woodland@curtin.edu.au

Joanne Ng
Client Facilitator
Library and Information Service
Curtin University of Technology
Joanne.Ng@curtin.edu.au

Abstract:

This paper discusses the analysis, design and implementation of an integrator system to share data between an institutional eprint repository (IR) and a University publications management system. The process of building IR functionality into the University system is described in the context of user demands to reduce data input to multiple systems, and the Library's desire to increase the uptake and usability of its own eprint system.

Introduction

The Open Archives Initiative and the trend towards eprint repositories and open access to research material

Since 2002, there has been a concerted effort in the scholarly communication field to developing a new model of providing access to publicly funded research, with material being accessible and available freely, in full-text on the Internet. The model is based on the principles and specifications of the Open Archive Initiative (OAI), an organisation and "movement" to "develop and promote interoperability standards that aim to facilitate the efficient dissemination of content"(Lagoze et al. 2002). The motivation to share and allow easy discovery and dissemination of research material spurred the development of the OAI-PMH - The Open Archives Initiative Protocol for Metadata Harvesting, which "provides an application-independent interoperability framework based on metadata harvesting" (Lagoze et al. 2004).

Thus started the trend for research institutions (particularly Universities) to create eprint repositories of their research output i.e. digital collections of research, housed in open source systems that were compliant with OAI-PMH protocols. The repositories are able to provide easy harvesting of their contents' metadata to search engines of academic material, and other harvesters e.g. Google. The full-text of the institutionally branded information is thus exposed to the international research community – without the need for end-user authentication, payment, proprietary software or other limitations. OAIster is an example of an academic search engine, currently holding 5,790,813 records from 527 institutions, and harvesting metadata from eprint repositories on a weekly basis.

Readings in the Open Archives field suggest that information managed in this way is more likely to be found, seen and potentially cited by other researchers, potentially increasing its citation rates and impact factors. The intended outcome is that the organisation's research is showcased and promoted such that it increases its research profile internationally.

Background at Curtin

Curtin University is a large organisation of over 30,000 students, approximately a third of whom are offshore and international. As a University of Technology it has a strong focus on applied knowledge, with a large number of undergraduate programmes. However, it also claims a strong research focus.

Curtin staff use two University systems to report on research activity and publication data. The "RPI" or Research Performance Index is a system that reports internally for the purpose of funding researchers. Funding is allocated for both inputs (in the form of grants and research income) and outputs (e.g. postgraduate research supervision and publications). Researchers who have demonstrated the potential to bring in external research funding for the University apply in RPI Groups for further internal funding.

The Publications system (PUB) collects information about the publications research output of individual researchers. It collects data relating to a pre-defined list of publication categories.

Research income and research publications data is provided to the Higher Education Research Data Collection (HERDC) process on an annual basis. The HERDC is conducted by the Federal Government's Department of Education Science and Training (DEST). Based on this information, the Federal Government determines its funding allocation to Australian universities. The process of data collection and reporting is a statutory requirement under the *Higher Education Funding Act 1988 (s18 (g)).*

Establishment of espace@Curtin

Deciding to be the first Western Australian university library to build an institutional eprint repository was something of a step into the unknown. In March 2003, repositories had a limited history internationally, and not surprisingly, a much smaller presence in the Australian library world. The Curtin University Library and Information Service undertook to provide leadership to the university in the area of scholarly publication by piloting a repository. Curtin's experience in the Australian Digital Theses Program (Council of Australian University Librarians 2005) provided useful background to the concept of a distributed database of digital content. However, the knowledge about how best to approach the management of eprints had to be built from the ground up.

Academic and research staff were consulted to gain some insight into perceived issues, potential problems and anticipated benefits.

The following factors were recorded as the group's primary concerns during these discussions:

- interoperability between DEST reporting systems and the archive
- ease of use
- ability to generate reports
- interoperability with EndNote bibliographic software
- the need for a simple model describing what can be submitted
- a means for recording usage and downloads i.e. statistics generation
- ability to extract information using XML
- ability to include unpublished or un-refereed material
- strict quality control.

Following endorsement from the University Research and Development committee (URDC) in December 2003 the pilot commenced, with the aims of building content to create a demonstrator repository model for the University, and a system which would serve the University's research needs and prove a useful initiative. An Institutional Repository Reference Group was formed, including some members of the URDC, to provide consultation and feedback during the pilot. One issue which continued to be raised was the desirability of having an eprint system which could share data with the University's current and future publications management systems.

Drivers to the integration project

There were several drivers behind the integration project. They include both significant external factors and local incentives from the education marketplace and Federal Government policy. One of the drivers, which have emerged most urgently in recent months, is the pressure on the Australian higher education sector.

The university is subject to market forces within an increasingly competitive international higher education industry. Falling domestic levels of enrolments are compounded by the increased need to attract international students, who have a wider choice of educational delivery programmes than at any other time. On a national level, the higher education sector is under review from political quarters which may place universities into less or more desirable groupings, affecting Government funding.

Overall, Australian educational institutions will receive $73.8 billion in funding over the next four years. In 2004 – 2005, $17 billion dollars is being received, an increase of 5.4% over the previous year's funding (Department of Education Science and Training 2004). A significant funding issue, apart from the increase of 34,000 student places, is the renewed focus on research and research infrastructure.

The Australian Government will be providing $542 million between 2004 and 2011 to provide researchers with:

- access to major infrastructure

- infrastructure funding more directly linked to Australia's National Research Priorities

- greater research collaboration and the collaborative use of infrastructure.

Of particular importance are:

- communications and information technology – especially high bandwidth and optical networks, and large-scale computing capacity

- management of and access to research information with the aim of maximizing returns on the national investment into research facilities and programmes.

The Federal Government's development of the Research Quality Framework (RQF) (Department of Education Science and Training 2005) to assess the quality and impact of publicly funded research has placed extra pressure on the University's conduct and assessment of their research activities. The RQF, when implemented nationally in 2007, is likely to be the basis of future research funding allocations. The ongoing pressure of an RQF process will force the University to examine the nature, management and impact of its research output – including publications data.

Locally, the citation rates and research impact of staff research is often not as considerable as it could be in a research institution. The University's research output is not always represented in a cohesive or easily accessible manner. Rather, it tends to be scattered across the organisation's Divisional web pages.

It is perhaps not surprising under these circumstances that the proposal for an eprint repository was favorably received. A system which promised to maximise access to, and exposure of individual and group research output, with the likelihood of increasing impact and citation rates, was largely seen in a positive light, including the contribution of *espace* data to the ARROW project (The ARROW Consortium 2005). Even those who lacked the time and initial motivation to participate by depositing material expressed an interest in contributing in the longer term. However, there was the recurrent complaint, or wish, that data could be shared between those systems that the staff were already required to use.

The management of the University's own publications data is fundamental to these issues.

Intended outcomes of the integration

The espace@Curtin system (*espace)* shares a common challenge with other repositories – building content and encouraging active deposits from academic staff. Interested individuals continue to contribute, but the uptake from the academic community is low at best. The library's willingness during the pilot phase to process material on behalf of academics inevitably resulted in older "backsets" of publications being offered for inclusion. On one hand, this demonstrated the important archival benefit that repositories provide. "Historical" research that previously existed only in printed form could be digitised and its access maximised over the long term.
For example: http://espace.lis.curtin.edu.au/archive/00000302/

However, it is generally accepted that citation rates are highest during the period immediately following publication. The inclusion of more in-press or recently published material in the repository would be likely to increase the potential benefits to Curtin researchers. By removing obstacles such as duplicate data entry, and by encouraging the deposit of new material, we hoped to increase the uptake and value of the *espace* within the Curtin research community.

The old publications system

The University's web-based administrative publication data system is managed by the Office of Research and Development (OR&D). The previous version had a general reputation for being difficult to use, unreliable and limited in its application.

Primary criticisms were as follows:

- data needed to be entered into two databases for both HERDC reporting, and for RPI reporting (for the purpose of internal funding). This applied an administrative overhead to individuals or departments which is resented by users

- users claimed that data previously entered has sometimes been "lost" when attempting to output the data at a later stage

- the data was limited to bibliographic details only, with little authority control or look-up tables to ensure consistency or correctness. Thus the data frequently exhibited typological mistakes and duplicated records

- there was no facility for the inclusion of full text documentation or digital objects - despite the data being related to the existence of these items, and their subsequent

processing dependant on the ability to view extended information about the original items

- little maintenance, and virtually no development, had been done on the system during its operation

- reporting from both systems was minimal, with output to basic Excel spreadsheets

- processes involved with verifying the data in the system required a manual, paper-based system with subsequent implications for records management

- the inefficiencies of the system repelled those users who could avoid the process. For example, users who attracted significant funding from alternative sources to the internal RPI process (e.g. from industry or other agencies) were not attracted to use the system for reporting research data to the University. They tended to enter data on personal or departmental web sites, bypassing the University's publication data system completely.

The new publications system (PUB)

The Office of R&D uses a variety of systems including Research Master ™ software.
When the Office commenced replacing the old publications database as a critical system, they required more flexibility than was available from the default ResearchMaster system. The decision was taken to build the new publications system outside this software. However, it was developed within the same software used for ResearchMaster reporting and for the University's Consultancy Coordination Unit (UCCU) and research grants. OR&D aimed to integrate the PUB system into the suite of OR&D databases, including those used for contracts and ethics. The architecture is Microsoft ASP.Net, with MSQL Database back end.

An opportunity

Aware that the old PUB system was due for replacement, the Library made contact with the OR&D. The window of opportunity in this case had been opened just the right amount. The OR&D were very interested in exploring the integration of the two systems, and the possibility of gaining library advice on additional PUB database functionality.

The profile of *espace* had been raised sufficiently during the previous year, both at Divisional and URDC level, to make integration a logical proposal. The approach was timely, given the pressure under which the OR&D were operating to replace the old publication system. The OR&D had commenced building the new publication "module", but it was in a sufficiently embryonic state to include all the repository features that would be required.

Choice of Integration Techniques

A major driving force behind this project is to achieve synergy such that data entered into PUB can be replicated to *espace* seamlessly. To facilitate the flow of data from PUB to *espace*, several integration methods were considered:

- Web service [1]

- XML messaging

- Direct database query

The choice of technique used was based on:

- cost and complexity – does the benefit derived justify the cost and time involved in implementing the solution?

- flexibility – is the solution adaptable to include new data sources?

- time to market – can the solution be implemented within weeks and not months?

- specialised skill sets – does it need specialised knowledge to develop, operate and maintain solutions?

- seamless integration – does it require changes to existing IT architecture, systems and applications to incorporate the integration solution?

Web service specifications and implementations are efficient at making request/response style requests between networked systems. Ideally, *espace* could provide a web service to allow remote systems to submit eprints. However, at the time of implementing this project, there was no provision for web service support by the EPrint™ software. While it would have been possible to build a custom web service for this purpose, it was less attractive due to the time and effort involved. Implementing the web service outweigh the benefits to be derived from this approach. In addition, the web service option would be justifiable only if there are multiple external data sources that need to be integrated with *espace*.

The XML messaging option would involve building a middle messaging format where PUB and *espace* could interact and exchange information with each other. This option is ideal if there are bi-directional information exchanges between the systems. As the data flow in this integration is one-directional, i.e. from PUB to *espace*, it is deemed that there would be little benefit derived from investing into building a middle platform.

The direct database query option was chosen as it was less complicated and provides maximum value in terms of the effort involved. Furthermore, it did not require any changes to the existing IT infrastructure. A custom integration module, known as the *espace* integrator, was implemented to extract, transform and load data from PUB and then create the corresponding eprint object in *espace*. We also chose to adopt batch mode integration. The integrator runs outside system peak hours thus relieving network and system load on both systems. Currently, EPrint™ software provides two mechanisms for importing eprint records from external sources: XML or EPrint™ APIs. The latter approach was selected as it provides more flexibility and control and allows eprint documents to be uploaded to the archive database.

The relationship between the *espace* integrator and both the systems are illustrated below:
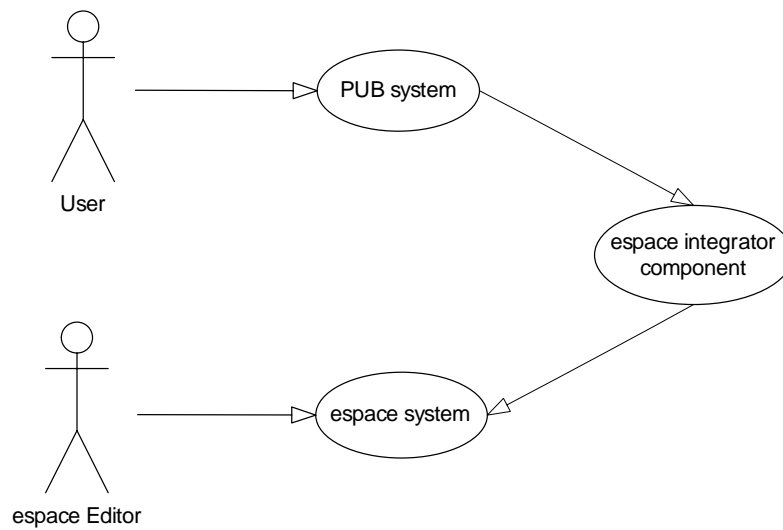

Figure 1: Relationship between Integrator and Existing Systems

Staff members who choose to use the PUB system will have the benefit of having their publication records replicated to *espace* automatically. On the other hand, staff members who wish to gain early exposure of their research findings, but do not necessarily wish to create the publication record in the PUB system, can continue to use *espace* as a self-archival repository. The data replication is one-directional, i.e. from PUB to *espace,* because PUB has a larger dataset compared to *espace*.

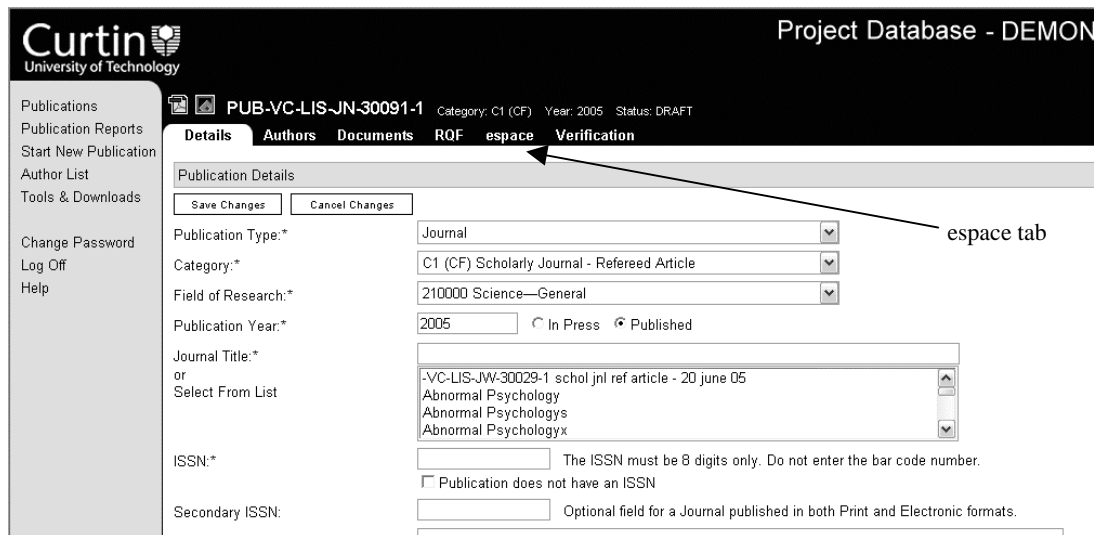The *espace* tab is presented alongside other PUB database options:

Figure 2: Web Interface of PUB System

Within the *espace* tab, the user is asked to

- choose deposit to *espace* (otherwise ignored by the integrator)

- include abstract, or tick "No abstract available", which populates a mandatory *espace* field with this text

- keywords

Figure 3: Additional espace Information in PUB System

Business Flow

Having decided on the integration technique to use, the end-to-end process flow of adding a new publication record via PUB to *espace* was mapped out. An overview of the process flow is illustrated below:
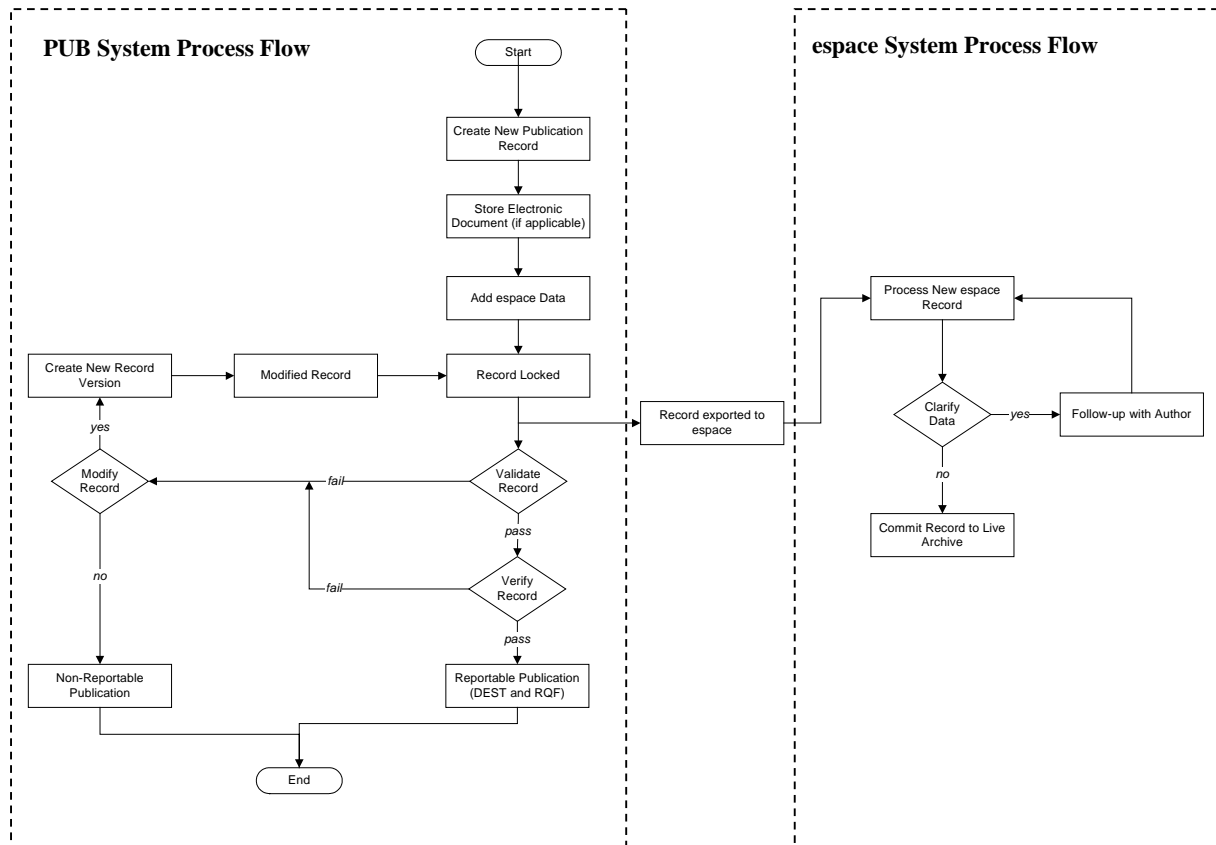

Figure 4: End-to-end Process Flow of Integration

(Office of Research & Development Curtin University of Technology 2005)

The user enters a publication record through PUB system, including *espace* data and the upload of an electronic copy of the document text. At this stage, the amount of *espace* data entry is kept to a minimum as the majority of fields can be mapped from existing PUB data.

Once the PUB record is entered and locked, the *espace* integrator will extract the record and populate the datasets into *espace*. The editorial process in *espace* will process the record and follow-up with the author for any missing data or to seek further clarification on the record. Once a record is verified by the *espace* editor, it is made available in the *espace* live archive.

Through past experience, it is found that about 10-20% of the PUB records might be later modified by the creator. This means that such records might potentially be extracted for export when the modifications have been entered and locked. To simplify the interface, the initial implementation phase does not take into account modified records. Instead, the *espace* integrator produces a nightly report on records that already exist in *espace* but have been modified in PUB. The *espace* editor will then use this report and decide if a follow-up is necessary to check with the author regarding amendments to be made to the eprint record in *espace*.

Implementation

Methodology

The implementation took about 4 weeks to complete and the steps adopted are as follows:

- gap analysis – determine the difference in record structure between PUB system and *espace*

- design and model – establish the end-to-end process flow of creating a new record in PUB system and populating it in *espace*, including the editorial process flow in *espace* in relation to the handling of these records

- mapping – map each data field in PUB to *espace* field

- prototype – develop a prototype to validate the process flow

- data cleansing – refine the data mapping rules and handle data exceptions

- coding and testing

The *espace* integrator was written in Perl to take advantage of the EPrint™ built-in APIs and comprises three components:

- extraction – probe for new data in PUB system

- translation – mapping PUB data structure to *espace* data structure

- load – creation of eprint record in *espace*

A separate module provides the audit trail of the ETL (extract-translate-load) process and this information is used in the editorial process in monitoring and processing these imported PUB records.

Managing the records received from the integrator

Within the EPrints™ software, a user must register in order to directly deposit material. The user has an identity and a User Area Homepage, with an "inbox" to start and edit draft records. This process was copied in order for the integrator to collate the extracted records into one area. "PUB ORD" was created with appropriate library authentication and the status of user. The espace@Curtin editor can authenticate using PUB ORD's details, and view records retrieved via the nightly extract. At this point, the records remain in the Submission Buffer – not yet live or accessible by the public. Checks can be made of bibliographic details and the nature of the attached full-text document, and the necessary permissions obtained in order to make the record online in the main archive.

Security

Security of the system is enforced using firewall rules and database privileges. The *espace* server resides behind the library firewall and rules are enforced to only allow specific network protocols between the PUB system and *espace* server. In addition, database privileges are defined at the PUB end to only allow specified IP and user accounts to probe the data.

Challenges

The data integration of PUB to *espace* presented a number of challenges during the course of project implementation. They are:

- incomplete data

  Not all data required by *espace* is captured in the PUB system, for example, the paper's abstract. If assumed that an abstract is not available, the *espace* editorial process would include an additional step to copy the abstract from the attached document or obtain it directly from the author. The objective is to keep the number of data entry fields to a minimum, and thus all mandatory data required by *espace* are solicited upfront from the user on the PUB system.

- timing

  A PUB record undergoes four distinct phases: draft, entered, validated and verified. The natural progression is from draft → entered → validated → verified. The time lapse between the draft and verified stage could be as long as 6 to 12 months. If the *espace* integrator were to probe data only in the verified stage, then it would lose the benefit of leveraging *espace* to gain early exposure in the research community. On the other hand, if data is probed at an early stage, the record may be subjected to frequent changes which make the *espace* update process too complicated. Upon further analysis of past records activity, a decision was made to probe data at the entered stage as this stage provides a higher degree of record stability and the time lapse between draft and entered stage is relatively close.

- duplicated records

  It is possible for a PUB record to be demoted from a validated to draft stage due to data error. This means that when the record is subsequently promoted to entered stage, the *espace* integrator will probe and create the same record in *espace*. While the eprint system allows multiple versions of the same eprint record to co-exist, we have made a conscious effort to keep the number of eprint versions to a minimum to avoid confusing users when searching and viewing eprint documents directly from *espace*. In most instances, the data changes that happen are mainly related to PUB record structure.

- data mapping

  The data definitions in the PUB system are quite different from those used in *espace*. Therefore, considerable time was spent in analyzing and mapping individual record fields in the PUB system to *espace*. For example, the PUB system has 29 publication types while *espace* provides 11 types. We did not attempt to expand the *espace* types to accommodate the wide range of publication types available in PUB as the *espace* types are widely understood and used in the eprint community.

  As the PUB system is essentially an administrative database for reporting to DEST, data translation was necessary to map ORD values to *espace* expected values. This is exemplified in the mapping of DEST categories, which are recognisable in the column below for PUB Publication Type.

espace@Curtin Publication Type	PUB Publication Type
Journal (paginated)	C1 – Scholarly Journal – Refereed Article
	C2 – Scholarly Journal – Other Refereed Contribution
	C3 – Scholarly/Professional Journal – Non-Refereed Article
	C4 – Journal Article – Non-Refereed Letter/Note
	C5 – Journal Article – Invited Papers
Journal (online/unpaginated)	Not applicable
Conference Paper	E1 – Conference Publications – Full written paper – refereed
	E2 – Conference Publications – Full written paper – non-refereed
	E3 – Conference Publications – Extract of paper
Conference Proceedings	E4 – Conference Publications – Edited volume of proceedings
	E5 – Conference Publications – Keynote address
Department Technical Paper	Not applicable
Department Working Paper	K – Reports
Newspaper/Magazine Article	Not applicable
Preprint	Not applicable
Book Chapter	B1 – Book Chapter – Of Book – Author Research Quality
	B2 – Book Chapter – Other
Book	A1 – Book – Authored – Research
	A2 – Book – Authored – Other
	A3 – Book – Edited
	A4 – Book – Revision / New Edition
Other	J1 – Major Creative Works
	J2 – Minor Creative Works
	J3 – Individual exhibition of creative works
	J4 – Group exhibition of a creative work
	D – Major Review
	F – Audio Visual Recordings
	G – Computer Software
	H – Refereed Design
	I – Patents
	L – Curating an exhibition
	M1 – New Media Creation
	M – Other

Related issues

Other issues that required consideration during the process were:

- retrospective data

    Did we want legacy records from the old PUB system? A decision was made to ignore the previous four years of data on the basis of poor data quality, the lack of resources to process such a large dataset, and the *espace* focus of early exposure of current material. We started with a clean slate and elective deposits to *espace* from 2005.

- full-text documents

    espace requires the inclusion of an attached document as a part of the deposit process. Once this specification was outlined, the OR&D appreciated the benefits of having an electronic copy of the document attached to the PUB record by default. This assisted the OR&D verification process as required by DEST.

- versions of document

    The system asks for deposit of the "Author's Final Draft – as accepted for publication, post refereed where applicable" as the preferred version. In line with the policy of many publishers to permit posting this version of an article, obtaining the author's version initially will streamline the editorial process. Publisher's versions and PDFs, for example, can be obtained easily via databases should the need arise. Multiple documents can be attached in PUB. Attaching a document is mandatory for *espace* deposit, but optional for PUB. The Documents tab offers two options:

- author's final draft
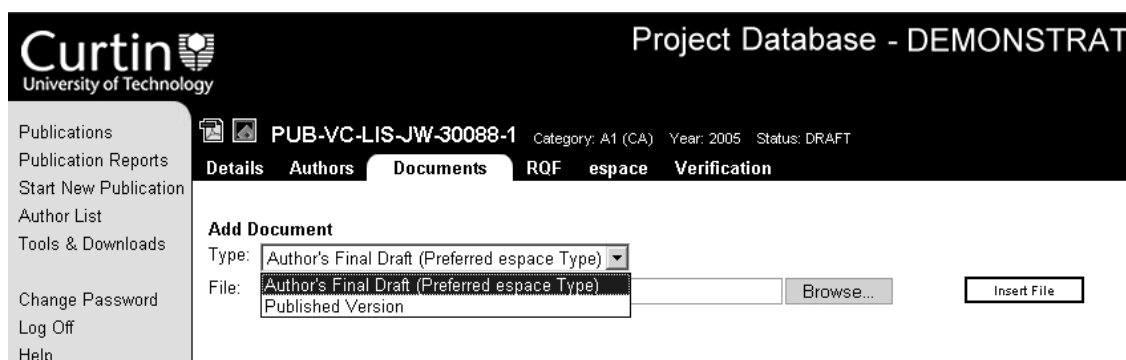
- published version

Figure 5: Document Types in PUB System

- file formats

  The EPrints ™ software specifies that a valid document type must be attached to a record in order to accept the file.
  The valid file types were:

File extension	eprint type
.pdf	pdf
.html	html
.htm	html
.ps	ps
.ppt	PowerPoint
.doc	Word
.rtf	Word
.txt	ascii
.text	ascii

  Any other file type would have been mapped to Other.
  If a file type of Other was attached, the software required the additional submission of a recognised file type. The PUB system had the potential to present the integrator with a variety of formats, including those which fell outside the EPrints ™ requirements, this halting the submission process.

  The script was modified to allow the type of Other to go through the submission process without having to attach an additional valid file type, thus preventing use confusion and downstream processing

- suitability of content

  The PUB system will accept all DEST categories, but would *espace* require only a smaller subset i.e. refereed material, according to its "collection policy"? We removed the decision from the user side, and accept all material, with secondary "filtering" for material suitability based on other *espace* criteria as a part of the editorial process.

Adaptation on the PUB side

Configuration and changes were not all necessarily on the *espace* side.

In a number of cases, the PUB configuration was adapted to provide better and more bibliographic data than had been available in its ageing predecessor. The OR&D were happy to incorporate changes to bibliographic fields, including volume and issue information, and the capacity for e-ISSNs and 13-digit ISBNs.

Populating a mandatory Subject field in *espace* was achieved by mapping from the PUB Field of Research value. After clarifying with DEST as to the minimum granularity required when using the ASRC schema, the OR&D adapted their "subject" list to match the existing Australian Standard Research Classification (ASRC) (Australian Bureau of Statistics 1998) headings used in espace@Curtin. Limited to the highest level of the ASRC schema, the mandatory PUB field records map neatly to the more detailed set of *espace* subject tags.

Fortunately, the EPrint ™ software deliberately keeps its number of mandatory fields to a minimum. All possible pre-existing, relevant PUB fields were mapped to *espace*, whether mandatory or not, in order to provide a better level of metadata in the repository.

Thus the analysis and planning was very much a two-way process between the developers, extending to library advice on PUB functionality such as EndNote exporting and formats for bibliographic reporting.

Reporting

Monthly reports are run on the following:

- number of record extracted from PUB ORD
- audit trail of extracted records
- duplicate eprint titles (where users may have added from PUB and via direct deposit)

Conclusion

The analysis, design and implementation of the integrator system have been completed largely without problems, and within the scheduled timeframe. Detailed analysis and ongoing communication between the Library and the Office of Research & Development was a central feature of the implementation. The system was launched in July 2005 to University staff. At the time of writing about 150 papers had been entered into the new PUB database, of which only a very small number had been "finalised", had their data entry completed and become available to the integrator. It is anticipated that, consistent with previous experience, the traffic will become more congested toward the end of the next funding cycle, when researchers and their administrative staff enter a large amount of data. The small number of records which have been extracted by *espace* appear to have worked successfully, with the end-to-end flow matching specifications. The process will be monitored closely in the coming months to ensure that the integrator meets the demands for user convenience, and continues embedding the espace@Curtin system into the University's research culture.

References

Australian Bureau of Statistics 1998, *1297.0 Australian Standard Research Classification (ASRC)*. from http://www.abs.gov.au/ausstats/abs@.nsf/0/51C2BDB99EBA43E8CA256889001E9D9E?Open&Highlight=0,rfcd.

Council of Australian University Librarians 2005, *Australian Digital Theses Program*, Council of Australian University Librarians. Retrieved: 15th December 2005, from http://adt.caul.edu.au/.

Department of Education Science and Training 2004, *Australian Government Department of Education, Science and Training Budget Information 2004 At a Glance*, Commonwealth of Australia, Canberra. Retrieved: 25 July 2004, from http://www.dest.gov.au/at_a_glance.htm#highered.

Department of Education Science and Training 2005, *Research Quality Framework*. Retrieved: 15 December 2005, from http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/research_quality_framework/.

Lagoze, C, Van de Sompel, H, Nelson, M & Warner, S 2002, *Open Archives Initiative Frequently Asked Questions (FAQ)*. Retrieved: 17 June 2005, from http://www.openarchives.org/documents/FAQ.html.

Lagoze, C, Van de Sompel, H, Nelson, M & Warner, S 2004, *The Open Archives Initiative Protocol for Metadata Harvesting*. Retrieved: 17 June 2005, from http://www.openarchives.org/OAI/openarchivesprotocol.htm#Introduction.

Office of Research & Development Curtin University of Technology 2005, 'Publications Flow Chart'. Curtin University of Technology, Perth, 7th September 2005.

The ARROW Consortium 2005, *ARROW : Australian Research Repositories Online to the World*. Retrieved: 9 September 2005, from http://arrow.edu.au/.

Notes

[1] The term Web service is used here to refer to a modular application that can be invoked through the Internet. The consumers of Web services are other computer applications that communicate, usually over HTTP, using XML standards including SOAP, WSDL, and UDDI.