

Managing digitisation of course readers through internal and external partnerships

Stephen Parnell*, Mike Hobart* and Brent Wilson ‡

* University of South Australia

stephen.parnell@unisa.edu.au, michael.hobart@unisa.edu.au

‡ Fuji Xerox, Australia

Brent.Wilson@aus.xerox.com

As is the case elsewhere, the University of South Australia Library scans printed material for inclusion in its electronic reserve collection and creates records for these as part of ensuring their availability for its staff and students. Where digitisation here differs from that of most other libraries however is in the close integration between the processes of two separate units within the University to make available electronically the content of printed readers. In this case, rather than scanning from original source material or from the printed readers, the Library-based Digital Resource Management Centre draws upon print masters produced within the University's Document Services Unit as part of the process of printing readers using Fuji Xerox Digipath technology. FlowPort, another Xerox product, provides a complementary avenue for creating digitised readings. The content of all digital readers is housed in a Library repository and made available through the catalogue as well as through the University's online teaching and learning platform.

Introduction

The Digitisation Resource Management Centre (DRMC) was established by the University's Senior Management Group in July 2002 as a pilot project to create and record electronic copies of excerpts from books, journals and other print publications. The DRMC is a major component in ensuring University compliance with copyright law and the licence agreement between the AVCC and the Copyright Agency Limited (CAL). The Library manages the DRMC and digitised copies are stored in a central repository on the Library server and used to support online teaching and other academic programs. The Library catalogue has been designated by the University as the single reference point for authorising digital copies of works in print. Having a single reference point that functions as both catalogue and digitisation register avoids duplication in record keeping while taking advantage of the features already standard in library catalogues - authority control for forms of names and full citation information. Unlike the predecessor copyright register, the Library catalogue is quality assured and now widely known.

In common with most other Universities, the Library scans material for inclusion in its electronic reserve collection and creates the records and links that ensure this material can be made available to all students and staff of the University of South Australia. It also provides direct links to individual journals and articles held in the various databases and electronic journals subscribed to by the Library and referred to in courses. Links to these form an alternative to scanning. The purpose of the DRMC and the eReader program is twofold. It acts to convert printed course readers into a digital format and it provides a mechanism to provide access to these directly from within the University's online teaching platform UniSAnet as well as via the Library catalogue. In achieving this outcome the DRMC also provides a mechanism to register and manage digital copies communicated by the University, facilitating copyright compliance. Digitisation at the University of South Australia differs from many other Library-based digitisation facilities however in the mechanism by which it makes available the content of printed readers used to support the University's flexible

delivery programs. Rather than scanning from original source material or from the printed readers, in most cases the DRMC is able to draw upon print masters produced and stored within the University's Document Services Unit as part of the process of printing readers using Fuji Xerox Digipath technology. The same company's FlowPort software provides a complementary avenue for making digital copies of print originals.

Digipath, DocuShare and FlowPort

Xerox Digipath provides a comprehensive suite of hardware and software for the University's digital print production environment. It includes four primary software modules: Document Scan & Make-Ready, Document Library, Web Services, and Production Print Path. DigiPath's scanning, digital printing and publishing technology is increasingly used to replace short-run black and white and color offset printing. This is not particularly new technology and in the University of South Australia environment it utilises propriety hardware such as scanners and docutech printers. Scanned images for printing are stored in the Xerox RDO format which is converted to PDF for the DRMC and wider use.

DocuShare is a software-based tool that is primarily used at the University of South Australia (UniSA) as a repository for the files and information created as part of printery operations. This is a web-based document management system and as such it is possible to store information and share it across internal departments, vendors and customers. It's secure, flexible and simple to use and permissions can be configured for better searching and viewing rights. At UniSA it sits on the same server as FlowPort. The DRMC and the Bob Hawke Prime Ministerial Library are set up as separate collections within DocuShare.

FlowPort is a server software platform used within the University to facilitate paper document transmission as part of the digitisation process. It thus provides a bridge between the paper and digital worlds by enabling users to capture and integrate paper-based documents into the University printery's digital workflow where content can be accessed, retrieved and distributed onwards, and all without the use of a PC by requestor or recipient. This is possible because a number of networked Xerox copiers in operation within the University include enhanced functionality that enables them to send images directly into the DocuShare server located in Document Services. While documents are typically stored in Xerox propriety format, these can be stored as PDF and from here can be picked up by the DRMC for inclusion in the University's digital repository once they are recorded in the Library catalogue.

For University staff the big advantage of FlowPort is that it can be set up so they never actually need to use a computer while scanning documents that will be added to the electronic repository. In support of the University's digitisation program the FlowPort administrator within the Document Services unit has created a number of PaperWare Forms, which include a series of destinations such as email addresses or document repositories. These sheets are printed out, and then when scanning a document, the user ticks the relevant boxes and uses the form as a cover sheet for the document. FlowPort can then send the document to the destinations selected on the cover sheet. You can use the form to nominate to send to any number of different repositories. Additionally FlowPort supports OCR and can convert the documents into any text based format required. . This is part of the UniSA installation and in the case of the Bob Hawke Prime Ministerial Library an "invisible" text file is attached to the PDF file which aids in the discovery of content. The scanner hardware used is all Fuji Xerox.

While the full potential of FlowPort is yet to be fully investigated, there are a number of issues that militate against it replacing the existing Library - DocuShare process. Images are created through FlowPort at 300 dpi and this is fine for material that will only ever be viewed on the web. In contrast the DocuShare - print master process creates images at 600 dpi, essential if there is any likelihood of the material appearing in a printed reader in the future. The difference is particularly noticeable for graphics. There are also the overheads that must be considered if digitisation activities are largely decentralised. One of the key considerations is load on network. Testing has yet to be carried out on the effect of simultaneous high volume scanning and document transmission from multiple FlowPort stations across the University. One of the major attractions of the Library relying in

the first instance on retrieving digital items that are mostly created within Document Services as a by-product of the print process is that in this arrangement the images used are created by experienced staff using high speed, high capacity equipment with specialist functionality. As a result the material made available to students is consistently clearer and cleaner than that provided from sources without similar technology or quality assurance process.

Process overview

Requests are typically received by the DRMC via e-mail or telephone, although items may be received directly from academic staff as print or electronic copies. All requests for material to be digitised to support the teaching and learning activities of the University require completion of a mandatory citation field consistent with the requirements of the Copyright Amendment (Moral rights) Act 2000. This requirement is also part of the remote scanning Xerox FlowPort system introduced by the Document Services unit in 2002. The information required is that generally accepted as good academic practice for citation of sources used. On receipt of the request, individual articles and chapters are checked against the catalogue and if not already available digitally, an order is placed with the Document Services unit for the PDF of the individual readings. Where possible the request includes the original print request number as this is used within Document Services to identify masters. The number is incorporated in many of the printed readers or can be found on a master list made available to the DRMC. This number is to be embedded in the PDF made available to the DRMC via Xerox DocuShare. The content pages of readers are then checked for required bibliographic information and copyright compliance. Depending on workloads within the Document Services unit, readers may be loaded onto a shared drive within DocuShare as a single PDF or as separate readings. CD-ROM is used as a backup delivery mechanism. The DRMC creates individual records for each copyright compliant reading. Once records are created for a complete reader and the files stored, the requestor is sent a list of URLs. All queries related to readings during the checking process are referred back to the requestor.

ImageServer

The PDF for each reading is transferred from DocuShare to the Library server once they have been checked for compliance with copyright. A coversheet is added using Adobe Acrobat and the complete PDF transferred to the digital repository using ImageServer, an add-on module of the Voyager library system. The coversheet includes a statement on copyright, citation details and the Digipath reference number. ImageServer provides a single repository and efficient, easy storage and management of scanned documents. Its use means that it is easy to allocate documents to open or restricted digital collections. Changing a collection from unrestricted to restricted access and vice versa is simple, instant and reversible. Access to restricted collections is by Voyager patron logon. ImageServer provides cumulative counts of viewings at the item and collection level and it tracks first-available-for-use and any anniversary data. It is possible to link ImageServer tables to Voyager tables to create queries to populate CAL data collection spreadsheets. This functionality was used successfully during the external digital copy monitoring period for the University. As important for those requiring access to readers, because ImageServer is integrated with the Library's cataloguing module, it is possible to use templates including derivation templates for linking parent-child bibliographic records. This streamlines and simplifies copyright compliance checking and individual readings for a course eReader can be linked together so that students can find all the readings. As records are passed through ImageServer the URLs are 'sent to Cataloguing' to be added to either the bibliographic or MARC holding record and this saves data entry.

During the first four months of operation the DRMC digitised thirty-eight readers at the request of the lecturer or school responsible for the reader. A further seventeen were in process. This equated to 1172 individual readings from Course Readers. Even as this work was underway there were 6646 individual accesses recorded – and without publicity! While most readers have been made available at the request of lecturers, a number have been digitised on the basis of the number of students enrolled in a course as well as those supporting study at the University's Whyalla campus. The DRMC makes available 300 items each calendar month with each reading taking approximately thirty minutes to process. This covers verification of citation, creation or

amendment of a record in the catalogue and checking links. All records indicate the relationship between the digital component and the source (original book or journal) as well as the reader of which it forms a part. Since it is recognised that many students access this material over a standard telephone line, all records show the size of the file in kilobytes. Consideration is being given to including searchable abstracts for articles appearing in digital readers.

The relationship between Fuji Xerox, Document Services and the Library

Material submitted to the Library for inclusion in eReserve has been variable in the quality of the print copy. Poor originals result in poor digital copies. Original print resources scanned by Document Services as print masters for readers are manipulated to remove blemishes, black margins and adjusted for alignment. The software used is also efficient at reducing file sizes. For these reasons copies available through Document Services are preferred over those supplied by individuals.

At the time of the original discussions between the Library and the Document Services unit a number of issues were considered that might arise from close cooperation. While it was clear that an opportunity existed to minimise duplication of activities within the Library and save duplication across the University, it was apparent that successful cooperation in digitisation could well have an impact on the Document Services unit's income stream from printing. It was also considered likely that there would be reduced sales of readers to students, an activity conducted by a different unit of the University. It is too early to tell if that will be the case, but if it is, there could well be an impact on work loads and practices within that unit. Another area that requires investigation is the impact of widespread use of PDF across the University network and the implications for setting the size of student e-mail mailboxes and print quotas

Substitution of costs associated with printing and postage by those associated with digitisation was not considered in the original proposal but requires investigation as the unit originally responsible for the production and distribution of readers is withdrawing from this activity in line with the university's strategic move towards online teaching. That this is a general movement within the University has become clear even in the short period since the establishment of the DRMC. An increasing number of readers are now coming directly to DRMC for digitisation rather than to Document Services for printing with digitisation as a by-product.

There is a strategic relationship between the University of South Australia and Fuji Xerox. Under the terms of the Agreement the partners share knowledge and best practice to ensure the best return on investment for the University's Document Services unit. Part of the brief is to explore emerging technologies that will allow this unit to expand its range of services and expertise. As part of this strategic relationship, the University receives access not just to advanced digital document technologies, but to skilled operators and document technology experts. In return the University provides a test environment for untried technology and acts as a showcase facility for the higher education market. Scanning is a logical extension of this activity. Fuji Xerox provided the hardware and software for the digitisation and management of original material held by the University as part of the Bob Hawke Prime Ministerial Library. The experience gained through this activity illustrates the benefits of cooperation between the University and one of its industry partners and provided a sound basis for extending an existing and successful cooperative activity to include the Library.

Impact on academic staff

While the digitisation project has been reasonably successful from a technical point of view, it has raised some issues that are best characterised under the broad heading of teaching and learning. One of these is perceived loss of control over course content amongst some academic staff. This is most evident in instances where academic staff feel caught between encouragement to make readings available online and the constraints of copyright. There are also some differences in turnaround time between material put on eReserve and that made available through the eReader project. In both cases it is necessary to balance maintaining academic staff interest in submitting acceptable copies and documentation against the need for

accuracy and control.

There is increasing recognition of the strategic opportunities presented by having the same resources available to all students regardless of mode of study. Benefits include:

- The potential to move away from course reserve collections located in a physical library to integration within online course pages
- Links from within courses can be to articles at different levels—at the article level, to the journal or issue or to the Library or saved search strategy. It can also be to a collection equivalent to a printed course reader
- Readings previously distributed only to off campus students can be available to all UniSA students on campus or at home
- Opportunities to build in value-added services such as Print on demand (for example by Document Services) to meet needs of those disadvantaged by transition to digital readings and meet needs of students with disabilities
- Potential for integration with other digital resources including student work and theses
- Individual readings previously hidden from other students can be found by author, title, source and keyword through the Library catalogue

Of particular interest to academic staff in the context of teaching and learning is that the number of times a particular reading is viewed is recorded. This could assist in determining recommended readings when courses and course objectives are reviewed.

For the Library there are benefits in a reduction in loss or damage to materials through theft or vandalism but this has been more than offset by increased responsibility. With the Library now responsible for making readers available online, it has also needed to accept responsibility for reliable 24 by 7 access to the Library system for this controls access to online readings referred to through UniSAnet.

The relationship between the DRMC, eReserve and copyright

Apart from the DRMC, the library is also offering a service called eReserve. This allows lectures to recommend individual articles or chapters, which are not part of a reader, to be digitised and made available to their students. It also allows for individual lecture notes, exam papers and sample assignments to be digitised. While the workflow for the two services is similar, the checking and scanning of the items for eReserve are all carried out at the individual campuses, with only the record creation occurring centrally. DRMC staff determines the format and information required in eReserve records to ensure standardisation between this process and that for eReaders. Records for eReserve items are upgraded to eReader records where needed. DRMC staff also assist in the creation of eReserve records and assist with answers to questions about eReserve.

While this paper deals mainly with the relationship between the Library, Document Services and Fuji Xerox, there is a really a fourth partner, the University Copyright Office. In addition to their role of adviser to individual staff, the Copyright Officer encourages, supports and promotes the role of the DRMC in ensuring compliance with legislation. The DRMC provides a mechanism that allows staff wishing to make print material available online to check against a single register that the material is not already available in digital form and, in the case of book chapters, that their requirements comply with the Act. Current activities in digitisation do not alter the role of the Copyright Officer nor eliminate the need for Schools and individual staff to take responsibility for compliance with the Act.

The biggest issue for the DRMC has been the one chapter rule that permits only one chapter of a book to be digitised for the University as a whole. A number of readers processed to date, have included more than one chapter from a single book within

the individual readers. In some cases the individual readers have included more than 10% of the single book. Staff contact the lecturer responsible for the reader each time the limit is exceeded to request advice on which chapter is the most important to them. The time taken to obtain advice back severely delays the completion of the reader. A small number of readers started in August 2002 were unable to be completed until replacement readings were agreed upon in January 2003.

The digitisation project has been promoted through the University newspaper and on the Library website. DRMC staff have visited most academic Schools within the University and Document Services and Fuji Xerox undertook a complementary promotion of FlowPort to some of the academic divisions. At present the DRMC is working in collaboration with the University's Flexible Learning Centre to increase awareness of its work amongst the Schools.

Resolving competing demands

Course readers are generally processed in order of request and Schools indicate the priority order in which they require readers to be made available. Given that it is only permitted to communicate a single chapter from a book, there are frequent occasions when it is not possible to meet all demands. To date most requests for potentially infringing copies have been made by individual lecturers wishing to digitise more than a single chapter. The University has in place procedures to resolve competing demands from across different divisions. These reflect the University's needs as a whole, not just the individual lecturer's needs. At the request of the Teaching and Learning Committee, the Chair of the Copyright Committee convenes this group. The level of requests for digital copies does not yet exceed capacity of the DRMC but this is changing as the benefits of the service become apparent. The Centre will need to prioritise its time as competition between Schools for readers to be completed within the same time frame increases.

While the original purpose of the DRMC was to make available digital copy of existing readers the number of requests coming to Library to make available a digital copy only of new readers is increasing, and this has required consideration as to how such requests will be handled. While most requests to the DRMC for digitisation of existing readers still allow digital copies to be acquired as a by-product of the print process, it was recognised at the time the system was set-up that a move to digital only was a likely direction for Schools to take. Wherever possible, any collection of items received by the DRMC that is intended to support courses offered online are sent to Document Services for scanning rather than be scanned by the Library. Document Services provide this service on a contract basis. This is consistent with one of the key features of the eReader process where there are significant advantages to all parties in using Document Services for large volume scanning. They have the equipment and the expertise to clean-up scanned items, but more importantly, they have quality assured processes and do not need to rely on the whim of those who may only ever scan a single article or reader.

This is not of course to say that all issues have been resolved. One of the major existing limitations of the process is the lack of coordination within the University between printing and digitisation in terms of record keeping and consistency in copyright compliance checking. Document Services do not check with the DRMC before printing, they print readers on the assumption that content has been checked for copyright compliance by the requestor. Users make the reverse assumption, they assume that anything submitted for digitization and printing will be okay. There have been some unhappy clients once it is discovered that a printed reader has not been checked before printing. Such discovery will probably not occur until the requestor is notified by the DRMC that digitisation of certain component readings would result in infringement of copyright.

This is an issue of timing. Print copies often have short timelines – and the digital copy is most likely to be produced from the print master after checking by DRMC staff. Responsibility for the compliance of printed readers is decentralized and rests within the Schools. Digitisation is centralized and rests with the Library. There are a few other areas of minor incongruence to be addressed. It can for example take longer for the Library to receive files for Readers sent by Schools direct to Printery than for those sent via

the DRMC, and there are instances of Schools receiving advice that their printed reader is available before the converted file appears on the DocuShare server. One of the more unexpected issues that have arisen resulted from the DRMC selecting readers for digitization on the basis of high enrolment numbers of use of printed readers through eReserve. In this instance the academic responsible for the course requested that digitization not proceed on the grounds that he was arranging for the commercial publication of the reader.

Metadata and other developments

As discussed earlier the print ready versions of the readers are of a generally higher quality (600 dpi) than their web presented counterparts. Investigation is proceeding in the use of XMP (eXtensible Metadata Platform) to link the web presented resources stored in the Library's ImageServer database to the print ready resources in the Xerox Digipath database. The intent is to provide an easy way for users of the resources to request them to be centrally printed. This is seen as a way to provide copies that are cheaper and of higher quality than those reliant on local printing. This could also offer benefits in managing copyright compliance in the future.

XMP stores xml metadata within the PDF, and has Dublin Core as the metadata default schema. However, as the name suggests it is also eXtensible and this will allow us to add any property that we like. In our initial trials we have attached a property called "digipath reference" and so in this way our web resources reference their print ready counterpart. This same process could be used to embed DOI information within the PDF. Digital Object Identifiers uniquely identify any published resource and will be an important step into managing copyright, between as well as within institutions. Until now publishers and document supply companies have had difficulty enforcing copyright and managing rights in the digital environment. The British Library has worked with Elsevier Science to ensure the secure delivery of articles by integrating Adobe Content Server encryption and eBook Reader software into its electronic document delivery operations (British Library 2002). The University and Fuji Xerox believe a similar strategy can be used in a print-on-demand service.

One area that is being considered is OCR-ing documents as part of the scanning process for any new readers submitted. This is labour intensive if undertaken manually and time consuming if included as part of the process for documents submitted using FlowPort. It would however add value by making the text searchable. At the moment readers required by sight-impaired students are converted to OCR on a document-by-document (needs) basis and information on how to request this format is included in the catalogue record.

In this initial phase the DRMC's role is limited to scanning and managing records for printed works only. The vast bulk of activity remains associated with making digital copies of existing printed readers and this has been greatly facilitated through the ability to utilise electronic file already held within the DocuShare facility managed by the Document Services Unit. The Centre is considering issues associated with hosting or facilitating e-publishing of material produced by research students and the Document Services unit has indicated its interest in undertaking any retrospective scanning of these.

Conclusions

Work continues on improving workflows between the DRMC and Document Services to reduce the lead-time between ordering the PDF and its receipt. To date activities have been directed at managing university records for material digitised in support of teaching and learning to ensure compliance with copyright legislation, and quality assurance. There is clearly potential for the systematic digitisation of resources to lead to complete replacement of printed readers. The implications of this are still to be fully considered. FlowPort also has great potential, although as this is investigated further it looks to have greatest application as a mechanism by which academic units can deliver documents to the Document Services Unit for cleanup and later processing and communication via the DRMC.

The close cooperation between the Library, the Document Services Unit and Fuji Xerox has provided a mechanism to ensure that printed course readers are made available online in compliance with Australian copyright law and without unnecessary duplicated effort in scanning. Use of the catalogue as a record management tool has ensured that individual readings contained within course readers are also available to all UniSA students and can be located using multiple approaches.

References

1. British Library (2002). "Encryption technology heralds new era in electronic document delivery - The British Library, Adobe and Elsevier Science set standard for secure EDD": *Content*.
www.econtentmag.com/extra/2002/2002_1206/2002_1206_1.html