

## **Transforming collection description**

Colin Sweett

### **Abstract**

*The National Library of Australia has transformed collection description and access for its digital heritage collections. The fundamental change from an item-by-item process, to large-scale data collection achieves efficiencies, enables rapid access, and allows the collection of previously complex electronic formats from publishers, photographers and donors.*

*The award winning Digital Library Infrastructure Replacement Program (DLIR), completed in 2017, was built using in-house expertise in programming, business analysis and collection management. DLIR produced a cutting edge suite of applications and systems that gives the Library the ability to produce system generated information and reduce or eliminate duplicate processing activities. Staff resources can be deployed where their expertise is most effective.*

*The Library's collecting of born digital pictures and manuscripts now benefits from a simple online deposit workflow. Information flows directly from the creator to the consumer and donor created lists can quickly be converted and published as an online finding aid in EAD (Encoded Archival Description). This rapid turnaround of descriptive metadata not only improves access but also enhances relationships with donors.*

*Similarly, the edeposit service built to collect published material enables the collection of born digital publications. The service has exponentially increased the Library's capacity to collect while supporting publishers to comply with new legislation.*

*Rapid access to content is enabled as publisher supplied metadata, captured via the edeposit service, is automatically mapped to a basic catalogue record which enables the ingest and display of serial issues without any human intervention. This eliminates the most labour-intensive of library tasks; serial check-ins.*

*This paper will discuss how the applications and systems built as part of DLIR have transformed collection description and access for published and unpublished materials. By illustrating our innovative deposit services, we will show how by capitalising on publisher and donor metadata, the Library has revolutionised access to its digital collections, and has achieved significant processing efficiencies for collection management activities.*

## Introduction

The National Library of Australia is faced with the challenges of collecting, preserving and providing access to increasing digital collections from digital publishing and archives, referred to as the 'digital deluge'. It is the mission of the National Library to collect, preserve and provide access to Australia's documentary heritage. This encompasses both published and unpublished collections comprising books, serials, web pages, maps, music, pictures, manuscripts and ephemera. Much of our collecting is facilitated through the legal deposit provisions of the *Copyright Act 1968*, enabling the Library to collect a copy of works published in Australia. In 2016 legal deposit was extended to include digital publications. This presented a significant opportunity for the Library to engage in the collection of Australia's digital publishing.

The Library responded to these challenges by developing a bespoke digital collection management system, Digital Library Collections (DLC). Released in 2017, the system enables and delivers collecting efficiencies across the organisation in mass digitisation, ingest of born digital objects and delivery of digital objects.

In responding to the challenges of digital collecting the Library has developed a number of transformative projects, leveraging off the newly developed digital collection management system to enhance the description of our collections. In developing these projects, the Library adopted a number of principles which underpinned the work:

- 'Create once, use many times', if metadata is created in one system find methods to reuse. This led the Library to incorporate publisher and donor created metadata into our collection description;

- User centred design, creating systems with our users foremost in mind. This led to the creation of easy to use deposit options for publishers and donors;
- Large scale data collection, prioritising the collection of metadata and e-publications at scale to enable collecting efficiencies;
- Reducing redundant workflows, embracing digital collecting as an opportunity to review and enhance our existing workflows; and
- Focusing on staff expertise, providing opportunities for our expert staff to spend more time on 'value add' descriptive activities.

The four projects outlined in the paper demonstrate the Library's approach of transforming collection description. These principles were embedded in the design and implementation of these projects that enable the Library to meet the challenges of digital collecting, and have all transformed publisher or donor created metadata to enable efficiencies in collection processing and rapid access.

## **edeposit**

The edeposit service is a one-stop online service enabling Australian publishers to deposit Australian digital publications (books, serials, maps and music scores), their related cover images and publisher supplied metadata. The edeposit service launched in February 2016, as the new legal deposit provisions came into force, allowing for the collection of born digital publications. The edeposit service is a public interface of an end-to-end digital collection management system that is fully integrated with other Library systems. The Library's approach was to extract maximum operational efficiencies from the edeposit platform, repurposing publisher supplied metadata, and automating previously manual processes.

Publishers access the edeposit service through an easy to use web portal which provides step-by-step instructions on the deposit process. The Library chose to develop a portal as it represented the best user experience for publishers allowing them to create an account and access their deposit history. The portal also provides efficiencies for Library staff in keeping track of publisher deposits.

Publishers have two streams through which they can deposit their online publications:

1. Individual deposit for publishers via the web-based edeposit service; and
2. Bulk deposit for the transfer of a large volume of publications in a single transaction.

The process for individual deposit via the web-based interface is very simple: publishers deposit their publications and associated cover image via the web application. The system asks for contact details, as well as basic metadata about the publication, including title, creator and year of publication. The publisher specifies how they would like to provide access to their publication ranging from onsite to freely available. It was imperative to make the deposit process as simple, secure and effective as possible for publishers to encourage their compliance with the new legal deposit provisions.

Once the deposit is complete, the files and metadata are transferred seamlessly to the Library's DLC system where they can be viewed by staff. An automated weekly process transfers the publisher metadata into a MARC record which creates a basic record on the Voyager Library Management System incorporating Resource Description and Access (RDA) standards. At this point, the publication is available to the public via the catalogue and Trove. The Library made a conscious decision to

reuse publisher supplied metadata to create basic catalogue records as we deemed the quality sufficient to enable faster access to digital publications to Library and Trove users.

The simple reuse of publisher supplied metadata has a dual benefit of saving staff time upgrading the record, as well as providing a basic record almost immediately to enable access. As a legal deposit library, the National Library undertakes a significant amount of original cataloguing. In this context, the automation of publisher supplied metadata to create a brief catalogue record for e-publications represents a significant efficiency benefit.

edeposit facilitates broader access to electronic publications for the Australian community allowing publishers to choose open access to their publications, making these available through the Library and Trove. Non-commercial publishers can agree to a broader level of access than permitted under the *Copyright Act 1968*, with the potential to make a much higher proportion of the Library's current collection available to Australians. The proportion of publishers providing open access is significant with 64% of monographs and 57% of serials deposited so far allowing open access.

The development of the edeposit service was facilitated through a close working relationship between the Library's IT division and the collection area. The Library is fortunate to have an IT division to facilitate the bespoke build of the service. To ensure the service met the needs of the collection area Business Analysts were recruited from the area to assist IT in scoping and writing the requirements for the service. Throughout the development process staff from the collection area were involved in User Acceptance Testing. This testing had the dual benefits of improving

the system through feedback from staff and increasing staff familiarity with the service from both the user and publisher perspectives, resulting in a smooth implementation. The Library made a conscious choice to harness the expertise of its staff in both IT and the collection area to ensure the development of the edeposit service met the needs of both publishers and staff.

To establish the automated mapping process, the Library's IT team developed a mapping template from DLC, where the data is stored in fixed fields, to MARC format. The mapping process ensures that fields stored in DLC are mapped to MARC correctly. This mapping approach results from consultation between IT and cataloguing experts in collections management who, using their knowledge of cataloguing protocols, improved the standard mapping to ensure efficient processing of digital deposit materials. This collaboration is an example of the Library's approach to focus on staff expertise to enhance its services.

The implications of this mapping are that cataloguing staff are required only to upgrade and edit fields where they add intellectual value to a record. Investing time in subject headings, mapping to name authority records and correcting access points improves the discoverability of the content. Much of the laborious correction of MARC fields has been eliminated by the intelligent mapping of publisher supplied data.

The service enables innovations in the collection of serial issues eliminating the need for staff to check-in digital serial issues. edeposit uses publisher supplied issue data to identify individual issues for Library and Trove users. These issues are automatically sent to DLC where they are attached to the serial title record and immediately available to the end user. Serial issue check-in is one of the most labour

intensive activities and eliminating this for digital issues provides significant efficiency benefits for the Library and enables faster access to new digital issues for end users.

Many of the innovations in the edeposit service provide benefits to Library and Trove users and the broader library community, as edeposit provides timely access to digital publications to the Australian community. edeposit forms the foundation for the development of National edeposit (NED) and these inbuilt metadata efficiencies will soon be available to NED member libraries. NED will be discussed in more detail later in this paper.

### **Bulk deposit**

The alternative method of digital edeposit offered by the Library is bulk deposit, a negotiated deposit option for publishers with large outputs, which provides benefits to both publishers and the Library. The Library has identified Australian publishers with an annual output of over thirty monograph publications, multiple journal titles, or map sheets that may benefit from bulk deposit. Bulk deposit enables batch ingest of multiple electronic publications, together with their associated cover images and metadata. There are several methods of deposit either via FTP transfer or via third-party distribution systems such as CoreSource where the Library has established a 'channel'. The set-up process is by invitation with the Library approaching publishers offering to establish bulk transfer of their publications. Once contact is established, the Library seeks agreement with publishers on the access conditions they would like to apply to their publications. As the majority of publishers identified for bulk deposit are large commercial publishing houses, all have so far agreed to make their publications available onsite which allows read-only access.



Throughout the process, the Library assures publishers that it meets relevant IT security guidelines for the secure transfer, storage and access to their publications. There is a clear risk for commercial publishers in the misuse of their intellectual property and the Library ensures that the bulk deposit process and access to this material is in a secure environment. Establishing a 'distribution channel' through an industry recognised publisher distribution system CoreSource has assisted in winning the trust of publishers as they are already familiar with the system and its inbuilt security. The delivery system for digital publications is secure with the publication being delivered in a secure environment on networked PC's within the Library's Reading Rooms. These PC's restrict the download and copying of digital publications that are available onsite.

The Library arranges a test deposit to check the mapping and ensure the files provided are free of technical protection measures, such as Digital Rights Management (DRM). This test deposit is essential as the Library uses publisher supplied metadata in ONIX, XML or CSV formats to create metadata in the DLC system, which flows through to create MARC catalogue records for the e-publications. In the case of monograph deposits, there are publisher standards to implementing ONIX, however the Library finds that each publisher utilises ONIX slightly differently and changing practice between publishers can result in errors in the established mapping. Once the files are received and the mapping confirmed a larger deposit is arranged with the publishers. The Library seeks to obtain the digital backlist from publishers on initial deposit which in most cases includes e-publications published before the expansion of legal deposit in the *Copyright Act 1968*. This ensures the comprehensiveness of our digital holdings for Australian published material.

The bulk process has been very successful with a number of large publishers commencing deposit since the service became available in September 2016. These include Penguin Random House, HarperCollins, Pearson Australia, CSIRO Publishing and Hachette. Bulk deposit is also being successfully extended to serial issues and maps. Academic publishers Wiley, and more recently Sage, have deposited over 590 digital serial issues. Furthermore, the Queensland Government have deposited 40 000 map sheets, while Western Australia Landgate have deposited 3 000 maps with the Library.

As with the edeposit service, publisher supplied metadata is automatically ingested through DLC creating MARC records to facilitate immediate discovery and access to these publications. Publisher supplied ONIX records are a rich source of metadata used by publishers to supply information about an item to online booksellers. The Library mines this metadata to create rich catalogue records that include additional fields that enhance access. Primarily these include publishing industry subject schemas such as BIC, BIASC and Thema subject headings. These headings are automatically mapped from the ONIX through DLC to the MARC catalogue record. With over 6 000 bulk deposit monographs received since the service launched, there is now a significant dataset of publisher subject headings to facilitate improved access to this material.

The process for serial publications is very similar with the Library collecting packages of content from publishers' secure FTP sites. These packages consist of XML metadata, the PDF article and cover image for the journal issue. Tables of contents and other additions such as addendums are also included. The XML files contain the article text, abstract, and a reference to the journal title and issue for matching purposes. The Library decided to manage the complex hierarchy between serial title,

issue and article to improve the display of this content for the end user. The publisher supplied metadata creates a three-level hierarchy between the title, issue and article to facilitate the delivery of this content via Trove. This hierarchy allows for the delivery of a 'virtual issue' experience for the user closely resembling the experience of browsing the print serial issue.

The bulk depositing of maps entails similar complex hierarchies. Publishers submit map sheets and an index to the Library via FTP including a template CSV file with the accompanying metadata. These files are loaded into DLC with each map sheet creating a 'child' record with the publisher provided metadata that is attached to a 'parent' series record for the specific map series. A catalogue record is created for the map series and the individual sheets are linked to this main record. Using the hierarchical relationship maps sheets are delivered in Trove within their series order and accompanying index.

The benefits of bulk deposit are numerous to the Library, publishers, users, and the Australian library community. The clear benefit to both the Library and commercial publishers is the efficiencies in deposit and end-processing of material that bulk deposit affords. Publishers no longer need to box up large quantities of material for legal deposit but, at a touch of a button, can edeposit their publications with the Library using established digital distribution channels they already utilise to send their publications to online booksellers. Feedback from publishers has been positive, with comments such as the service is "quick and easy to work with when depositing".

The Library benefits from mining the richness of publisher ONIX, XML and CSV files to improve the data quality of edeposit records and provide additional access points through the use of publisher subject schema and article level metadata. The

Library's decision to facilitate the automated reuse of publisher metadata provides enhanced collection description for users that would otherwise entail significant staff time to implement. The efficiencies for the Library in end-processing are significant as much of the labour-intensive cataloguing work is already completed by publishers with the majority of bulk records requiring little, if any, manual intervention or upgrade. Library and Trove users benefit through timely access to records of bulk deposit publications. While the Australian library community benefits from this initiative through the availability of newly published content from Australian publishers and enhanced records that they can use when copy cataloguing items in their collections.

The Library will continue to explore innovative systems and workflows to improve both data quality and drive efficiencies. Methods to further automate the enhancement of bulk deposit are being pursued. Chief amongst these is the work to map ONIX creator records to the VIAF (Virtual International Authority File) enabling likely matches to be returned for cataloguer checking and enhancement. The automated querying of the VIAF authorities will achieve additional efficiencies in processing bulk deposit records and assist in improving the consistency of the Library and Australian Bibliographic Network authority files.

### **Prepublication Data Service**

The Prepublication Data Service (PDS) was developed to modernise the Cataloguing in Publication (CiP) service which been provided to publishers by the Library since the early 1970s. The CiP was a cataloguing service for libraries providing detailed prepublication catalogue records based on information received

from the publisher. This service had the benefits of providing libraries with cataloguing data and assisting with acquisitions activities, and enabling publishers to announce a forthcoming publication. Library staff would handcraft a detailed prepublication record based on information received from the publisher and the record exported to Libraries Australia and Trove. A review found that almost one third of CiP records created never eventuated in a deposit, thus representing a significant amount of staff effort expended on creating records for material that never eventuated in publication.

The Library decided that the CiP service must be modernised, but it should retain the benefits for publishers and libraries, such as providing a prepublication record for libraries to use for copy cataloguing and a notification to libraries and booksellers of an imminent publication, while reducing the burden on staff handcrafting records that never eventuated in publication. The key goal was to ensure that intensive cataloguing activity was completed once the item was in hand.

The CiP Transition Project was established in September 2016 to implement this work. The project consulted widely with publisher and library stakeholders and found there was support for the continued availability of catalogue records for new Australian books on Libraries Australia. The project then established an automated publisher notification service called the Prepublication Data Service. The scope of the PDS included many of the benefits of CiP via an automated process. The new service includes:

- publisher access to a 24/7 prepublication service, replacing the CiP application with an improved online application form;
- customer service point for publisher enquiries;

- legal deposit promotion for collection development;
- online distribution of prepublication metadata via Libraries Australia in MARC format, and
- automated system integration and information delivery of prepublication metadata.

Through the transition process, a new application form was devised for the PDS service. This included similar fields to the CiP application with the inclusion of publisher selected Thema subject headings. Thema is a global broad subject category scheme, which was selected for the PDS service as it provides straightforward subject choices for publishers and is being implemented by many Australian publishers. PDS incorporates subjects for fiction and non-fiction titles with publishers able to select up to two subjects for non-fiction titles.

One of the most significant changes during the transition to PDS was the implementation of a generic cataloguing statement to be printed on the imprint page. This replaced the CiP prepublication record that was previously provided to publishers. The cataloguing statement simply reads 'A catalogue record for this book is available from the National Library of Australia' accompanied by the National Library logo.

The PDS service is facilitated using the RefTracker enquiry management system which manages the online form and sends the data via a mapping system to the Library's ILMS. The ILMS team automatically include these records in the nightly export to Libraries Australia.

To facilitate stakeholder consultation and a smooth transition, the PDS was released in two stages. Many of the changes were implemented in the first phase in August 2017 to provide publishers with:

- Prepublication records generated automatically from publisher's application metadata;
- The ability to select Thema subjects for their upcoming book from a short list, and
- A cataloguing statement which replaced CiP entry for publishers to print on their imprint page.

During this phase, the cataloguing work is completed on the publication once the Library receives it. The second phase, implemented in November 2017, was the official launch of the PDS service with the official rebranding and updating of Library webpages.

The overall feedback from publishers has been positive. Most publishers appreciate the immediate turnaround for their applications, understanding that the information they submit will be transformed into a MARC record which appears in the Library catalogue and is exported to Libraries Australia and Trove. There has also been a positive response to the new cataloguing statement which is widely used by publishers. For the Library, significant efficiencies have been gained through modernising the CiP service. The PDS service requires half a full-time position to manage the publisher enquiry line and undertake limited corrections to records, whereas a team of four staff administered the CiP service.

## **Digital Deposit Service**

The Digital Deposit Service (DDS) is a web-based submission system used to invite and allow donors to transfer unpublished digital material, such as pictures and manuscripts, to the Library. Furthermore, it allows staff to manage and process these submissions in an automated fashion with extensive reuse of descriptive and technical metadata.

The service facilitates efficient management of digital archival material and the smooth ingest of donor material into the Library's DLC system. The DDS is an invitation only service ensuring that acquisitions staff have a conversation with the prospective donor before any submission, thus reducing offers that do not fit within the collection development policy. It allows Library staff to process digital donations with little movement and eliminates the risk in using email or portable carriers such as USB Flash Drives, which are prone to corruption.

The service can process a variety of common file formats taking into account the diversity of digital archival materials. The system is versatile, allowing single or multiple submissions per donor or from multiple donors. Files per submission are limited to a total of 5Gb and/or up to 1000 files. Once received the files are stored securely in the DLC system upon ingest ensuring their continued access and preservation.

The DDS system manages the complete acquisitions process, including the creation of descriptive metadata, and interacts with several internal Library and Archival Collections Management Systems, facilitating the transfer of technical and descriptive metadata between systems.



The system includes the following functions:

- Transfer of digital objects;
- Appraisal by Library staff;
- Description by donor;
- Ingest to safe storage;
- Rights management, and
- Reuse of metadata to create catalogue and finding aid descriptions.

Upon receipt of the submission, Library staff appraise the material within the DDS system. Staff can view each file with its associated donor description and accept or reject material on offer. Technical metadata is also reviewed to ensure any commissioned pictorial work meets Library standards. Once appraisal is complete, the system produces a report detailing what has been accepted and rejected for audit purposes. The accepted items are transferred to DLC for safe storage.

DDS enables the donor to submit descriptive metadata for the material. Adopting the principle of 'create once, use many times', this metadata is reused in assessment and description. The donor can either describe item by item or use a spreadsheet for bulk items. Once accepted the donor generated descriptive metadata is transferred to DLC and imported to other library systems where it is reused to create catalogue records and collection finding aids providing access points to the digital material.

The DDS has realised an improvement in the processing time for new acquisitions of digital pictures and manuscripts materials. The reuse of donor created metadata during deposit has decreased processing time in the creation of catalogue records and finding aids, while providing the end user with detailed information on the collection from the donors themselves. The system has eliminated data security

issues and corrupt files by ensuring file integrity checks are undertaken automatically on upload by the donor. The invitation only approach ensures that donors have discussed their material with an acquisitions officer prior to ingest reducing the appraisal time and communication around rejected submissions. The DDS enables the donor to complete a rights management (online) that specifies rights conditions for the collection of individual files or objects. These processing improvements result in timely access to these collections subject to the conditions of the rights agreement.

## **Conclusion**

The National Library's innovative use of metadata to transform collection description has realised collection processing benefits that translate into faster access to the national collection and internal productivity benefits.

DLC is now underpinning the development of National edeposit (NED): a collaboration between the NSLA libraries to create a coordinated national electronic legal deposit system. Due to be released in 2019, NED will transform digital deposit for publishers with a single deposit system to fulfil their legal deposit requirements for both state and national libraries. NED incorporates the key features of edeposit and bulk deposit providing publishers with the flexibility to deposit publications individually or en masse. NED will extend these innovations to the Australian library community providing immediate access to digital publications and productivity benefits to the state and territory libraries.

The Library is also pursuing opportunities to develop further our pre-publication service. There is a real opportunity for the Library to identify ways that the flow of

metadata between this service and the National edeposit service can be made more efficient for publishers. This will support work to streamline the identification and description of Australian publications.

The Digital Deposit Service has the potential to expand to support the collection and description of unpublished material in other Library collections, including ephemera and maps, to accommodate complex digital objects and increase transfer capacity.

This paper has demonstrated the benefits that can be achieved through transforming collection description. By embracing these opportunities, the Library has enhanced access to digital material (both published and unpublished) for users and the Australian community. The Library has clearly achieved its ambitions to respond to the 'digital deluge' by fostering the development of innovative and adaptable services prioritising the reuse of publisher and donor metadata to enable the transformation of collection description for the Australian library community. The Library continues to meet the challenges of modernising our systems and is soon to embark on a project to implement a new system for cataloguing and description, which will provide exciting opportunities to further transform collection description.