Development of a rights management system for the National Library of Australia's collections

Sandra Henderson
Executive Support Branch
National Library of Australia
shenders@nla.gov.au

Matthew Walker
Collection Infrastructure
National Library of Australia
mwalker@nla.gov.au

Abstract:
This paper explains some of the work done during the Rights Management project at the National Library of Australia, looking at the background to the project, the expected benefits, the process of design and development and user engagement, and explains how some of the many challenges encountered to date have been met.

Background

Rights, including copyright, are not the main thing people using the National Library are interested in, and many users struggle to understand why so many obstacles are placed in their way when they want a copy of an old manuscript, or want to perform a piece of music from the collection. In these days of "it's on the web so I can use it", it can be difficult for our users to understand the issue of rights. It can also be difficult for our staff to know the intricacies of copyright legislation, or the history of the rights pertaining to particular items.

The National Library of Australia has many millions of individual collection items, many of which are in copyright either because they are relatively recent, or they are unpublished and therefore subject to perpetual copyright. Much of the Australian special collections material is unique, and frequently sought by users wishing to access, copy or publish items.

The National Library has acquired the rights for some items, but for most of our collection the rightsholders are outside the Library – they may be creators, agents, publishers, heirs, family members, and may be organisations or individuals. For some materials the rights are held by many rightsholders – a piece of printed music, for example, can have a composer, a lyricist, an arranger, an editor and an artist. Ensuring that the rights are protected is very important to us – rightsholders trust the Library to look after the materials and manage access to them in an appropriate manner. The ongoing development of many of the unique collections relies on the close relationships we have built over many years with creators and donors who see the National Library as the appropriate repository for safekeeping of valuable collections. Ignoring the wishes of rightsholders puts those relationships in jeopardy.

Until implementation of the Rights Management System (RMS), information about rights, rightsholders and permissions was stored in a bewildering array of paper folders, card files, paper or electronic official records, catalogue record notes, and in some cases, only in the memories of staff of special collection areas. A manuscripts collection, for example, may have restricted access imposed for a certain number of years by the compiler of the collection, a privacy restriction imposed by the Library on medical or financial records, an official restriction imposed on government records, and copyright restrictions for published items within the collection. In most cases the corpus of information was not easily shareable with other collection areas, and often took some time to locate.

The MARC record does have fields that can accommodate some information about key contributors to the creation of an item, access restrictions, and use and reproduction. It cannot readily accommodate information relating specifically to past and current rightsholders or copyright interpretations. A recent MARC discussion paper (California Digital Library/Library of Congress 2007) looked at the advisability of creating a new 5XX field to accommodate information relating to copyright. This remains under discussion. If such a field were to be added to the MARC record there would still be an issue for some libraries, including ours, about how to record rights other than copyright.

So what is the aim of the rights management project?

The aim was *not* to implement a system which would itself enforce restrictions on access and use, as many true digital RMS do, nor was it to record information only about our digital collections. The RMS is storing information about digital and non-digital collection materials, and in some cases will store item level information for material catalogued at a collection level.

The benefits for the Library and its staff include:
- Recording and sharing of information about rights for individual items or collections;
- Recording and sharing of information about particular rightsholders. These include contact details and notes about the Library's relationship with them;
- Recording of permissions granted to users;
- Recording of changes in rights policies over time, particularly where this has been stipulated in agreements with rightsholders;
- Automated determination of copyright status, using an algorithm based on the provisions of the Copyright Act;
- Savings in staff time in locating copies of agreements and permissions, through direct links to relevant documents in the Library's electronic records management system;
- Improved awareness of the importance of recording appropriate rights information in catalogue records;
- The routine recording of rights information by cataloguers, using standardised phrases which are more easily interpreted by both staff and the RMS;

Project history
A feasibility study into rights management for the special collections was undertaken in 1997, so this is a project which has had a long gestation. When the Library embarked on large-scale digitisation some years ago, a tender process in 2000 failed to reveal any products which offered a sufficiently affordable and low-risk solution to digital collections management. As a result, the Digital Collections Manager was built inhouse and continues to be developed. The decision was taken in 2001 to develop the RMS as a module of that Digital Collections Manager. However, and importantly, the intention has never been to limit the RMS to digital materials, which are a tiny fraction of our collections. The National Library has an active digitisation program, and is currently digitising around 10,000 collection items each year, but much of the collection will never be digitised.

Scoping the project
It was clear at the outset that the RMS would need to be able to store and maintain information about rightsholders, about policies referring to rights, and about users who had obtained specific permissions, and store information about our own collection items as well as items from other collections used for exhibition, publication or other purposes, when we need to be aware of restrictions on any use of these items for publicity or other purposes.

It was decided that migration of existing paper-based data was outside the scope of the rights management project. The volume of such data means this data migration will have to be a very long-term activity. This does not preclude individual areas scoping the task of data migration and using various means to accomplish some of

this work. Some collection areas were very keen to be able to make the fullest use possible of the functionality under development. The Pictures area entered rightsholder data from paper files into Excel spreadsheets so it could be loaded to the RMS at an appropriate time. As part of the second phase of the project, guidelines for importing this data into the RMS have been developed in the IT division.

The RMS has been developed as a staff-access-only resource. Privacy laws notwithstanding, many rightsholders would not want their contact details made available to the public, and staff want to be able to include notes about the ease, or otherwise, of dealing with particular rightsholders, or, for example, temporary contact changes.

User consultations

A series of meetings were held with individual business areas to outline the project and get a better idea of how each area handled rights information, and also to find out what data stores existed. Later meetings have been held to discuss issues, demonstrate progress and seek feedback. As functions have been completed, all stakeholder areas have been encouraged to be involved in testing. Key stakeholder areas were identified as Pictures, Manuscripts, Oral History, Maps and Music, who all acquire, process and provide access to special materials, and Publications and Exhibitions, who are major users of special collections materials. The general collections area has responsibility for material that in most cases has very clear copyright, such as published books, but has remained involved in the project, and provided considerable assistance and advice with examination of the Voyager catalogue and authority records.

The policies applicable to particular items change over time, so allowance has been made for the archiving of a history of policy changes for an item, and a decision was made that the most recent policy would be the one reflected in the catalogue record.

Design and development

This section looks at the Library's current systems, the systems architecture for the Rights Management System, and the process of developing the system.

Current systems

The RMS had to be designed to support rights recording for all collection items, both physical and digital. This was a difficult requirement to meet due to the existing cataloguing systems used by the Library.

ExLibris' Voyager (previously an Endeavor product), whose cataloguing client is configurable via templating, is used as the Library's Integrated Library Management System (ILMS). However, it was discovered early in the project that even if the extensive configuration required could be achieved, the proposed structured data for rights could not be supported using standard MARC fields, as explained in the opening section of the paper. This realisation left the project with the task of building a standalone system for managing the rights, which would have to integrate with the ILMS via its faithful Z39.50 interface. The Voyager system is, like many other commercial ILMS products, otherwise virtually impenetrable for interoperability with other systems.

Not all descriptive metadata is stored in the Library's bibliographic catalogue. Many parts of the collection are only described at collection level, particularly in the case of manuscript collections. For items that are contained within these largely undescribed collections, the only accessible electronic source of descriptive metadata at the item level is the Library's Digital Collections Manager (DCM) system. The DCM, first released in 2001, manages the selective digitisation workflow for visual and audio materials, and contains the descriptive and other metadata for the digital items that have been digitised from the Library's physical collections. Since the DCM already integrates with the Library's catalogue, it was the next option considered as a basis for the RMS. This was an especially strong consideration because many of the proposed users of the RMS are already users of the DCM.

The Rights Management System
Since the DCM software was first created over seven years ago, it was implemented with what are, now, outdated technologies including an internally-developed Java web application framework. The technical staff allocated to the Rights Management project are not familiar with the old technology, and the Library has identified other new technologies which are preferable to use in the implementation of new software systems. This posed an issue because the most desirable option from a business perspective was to implement the rights management functionality as part of an existing operational system. The main reason was to avoid workflow inefficiencies caused by users switching between various systems as part of their everyday work.

The Library came up with an elegant solution in which the existing DCM software could be left virtually untouched, and the RMS could be built as a new technology "wrapper" around the DCM software. This solution is the manner in which the RMS has been successfully implemented. Users access the rights management functionality via the DCM user interface without noticing any difference in the legacy DCM functionality.

Of course, the Library also considered that it may want other software systems to access, create and update rights information. To make this possible, much of the core rights management functionality has been implemented in a modular fashion using Web Services technology in a Service Oriented Architecture. The Web Services implemented to date include services for: retrieving a bibliographic record in MARC XML format from the Voyager catalogue; determining the copyright status of a given item; and adding, updating and retrieving rights information from the rights management database.

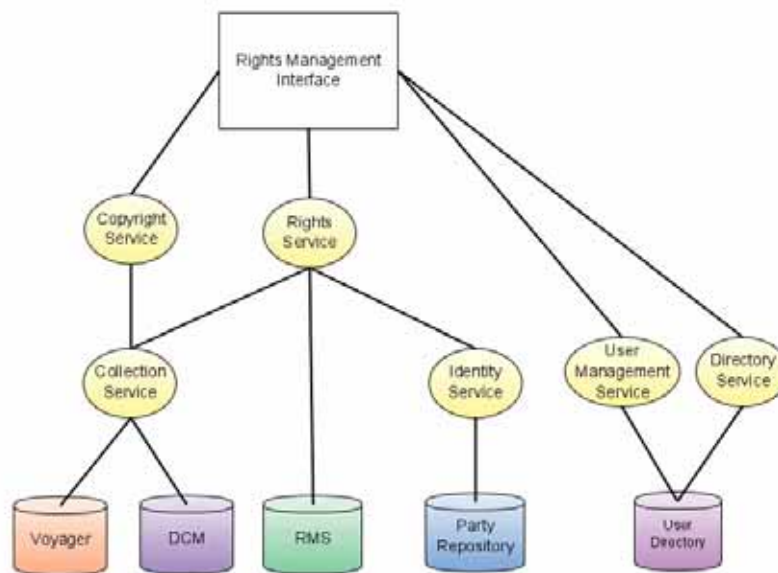Figure 1 shows the various components involved in the RMS.

Figure 1. Components of the Rights Management service.

All of these Web Services may be of use to other institutions interested in the maintenance of rights information. Since the Copyright Status service is based on Australian copyright legislation, it is likely to be of interest mainly to other Australian institutions. The Library intends to make all of these services openly available to other organisations. However, in order to use them, the other institutions would need to develop their own user workflow system to integrate with the services. The Library also intends to publish the copyright status rules documentation in diagrammatic form. These could also be of significant interest for those institutions forced to continue manual checking of copyright status.

In the first release of the RMS software, users are only able to access and record rights information for items in the digital collections, due to the fact that the DCM software only allows searching on digitised items. In the future, a search service will be implemented to enable efficient searching across the metadata available for both physical and digital collections. This work is planned for the third stage of the project.

The second stage includes integration with the Library's party database which, at the time of writing, is under development in the People Australia project. The party database is a repository for storing people and organisation information using the Encoded Archival Context schema. The People Australia project is developing a Web Service that will enable the RMS to create and access information about people and organisations with a rights holder role for collection items. Initial data will be derived from sources such as the Australian Name Authority File. This is an interesting aspect of the project, in that the separate directions and resources allocated to each project have had to come together to collaborate on building a system that will support the requirements of both projects. This is not something that came naturally – both project team members and management have struggled to come to terms with the overlap in requirements for the two projects. The situation

has brought one of the central issues of the IT Architecture Project Report (National Library of Australia 2007), the single business approach, to the forefront. As a result the Library has become more aware of the some of the issues surrounding the single business approach, avoided the duplication of effort in separate projects, and set a precedent for future collaboration in priority projects.

The People Australia project had been evaluating Encoded Archival Context (EAC) as the main candidate for storing the people and organisation data, and had identified some limitations that needed to be addressed locally and raised as issues for the long-term implementation of EAC as a standard. The Rights Management project team investigated how rightsholder data could be stored using the EAC schema, and came to the conclusion that the data structures were overly complex for project requirements. Both teams were in agreement on the limitations of the EAC schema. However, they agreed the limitations could be circumvented in a local implementation and that the Library could influence the future direction of the schema as a proposed standard.

The People Australia project team agreed to implement a Web Services interface to an EAC-based data repository that both projects could use in their implementation of user functionality. This development is still progressing and as at January 2008 it is impossible to state the success of the approach. However, the Library is confident that the approach will work, whatever issues may arise during the implementation.

Earlier in the project, a decision had to be reached about the data model for storing rights information, including policies for access and use of collection items as negotiated with the rightsholders. After experiencing difficulties with defining a simple data model for this purpose, the project team looked elsewhere for existing standards that could be used or adapted for this purpose.

The INDECS (INteroperability of Data in E-Commerce Systems) schema is the result of an international effort to create metadata standards for e-commerce in relation to rights owners. The Library's informal evaluation of it concluded that this schema's strengths are the recording of events and relationships that occur between parties (e.g. publishers, rightsholders) and creations (i.e. publications, collection items), but that it does not cater for the recording of rules-based policies that represent the control of access to creations, which is one of the RMS project's main aims. It would potentially have been useful as a schema for storing people and organisation data, if the Library had not made a decision to use EAC.

The eXtensible Access Control Markup Language (XACML) standard was investigated and selected as the basis for the data model to store rights information. At first, the project team considered using the XACML schema itself as the data model. However, after further consideration the consensus was that it was more complex than necessary to support the requirements of the RMS. While it may be useful as a data interchange schema, it was not necessary to incorporate that level of complexity internally. A decision was made to implement a simple data model that supported the basic requirements, and that mapped directly to the XACML schema through a simple transformation. That simple data model is now the basis for the first implementation of the RMS. Figure 2 shows a representation of the data model that has been implemented.
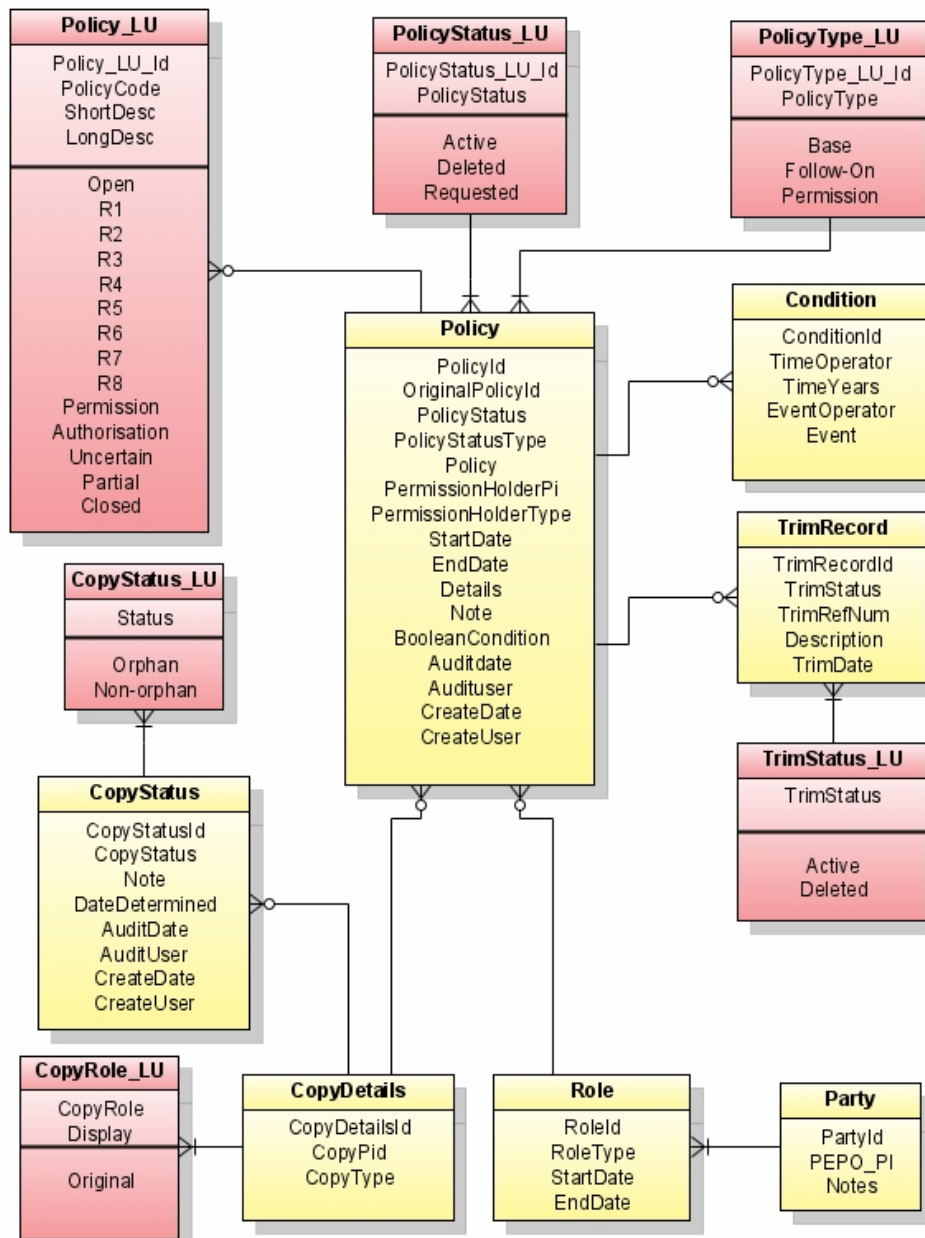
Figure 2: RMS Data Model 1.0

The issue of adequately but efficiently representing rights in catalogue notes for the information of users and also for migration to the RMS has been examined. Approximately 100,000 records in our catalogue contain some information related to access and use. One task undertaken was to examine this data and reduce the key information to a series of succinct statements which could be used by cataloguers in the future, and which the RMS can interpret as generic base policies. The statements include:

- Open (for those materials to which no restrictions apply);
- Closed (for some recordings and manuscripts which have a lengthy period of restriction); and

- A graded set of statements expressing varying availability of material for reference and reproduction purposes. A series of codes, R1-R8, is used in the above diagram to represent these generic policies.

As mentioned previously, the third stage of the project will involve the development of functionality to allow searching across both physical and digital collection items. This will, of course, allow rights information to be associated with any collection item. The details of this stage of the project are still being planned. However, at this stage the intention is to use Lucene/Solr technologies to build a search database containing the metadata for all described collection items, including linkages between item-level records for digitised collections and their containing collection-level records in the Library's ILMS. This search database will be built with other purposes in mind, and as a Web Service to ensure its usefulness for future initiatives. This will also enable the search database for the RMS, and the DCM, to be replaced easily should the need arise. For example, the Library is also working on building a search service including data from all of its sources, including the ILMS and the National Bibliographic Database (NBD).

Challenges

Originally planned to be a project lasting a few months, the RMS project has proved much more challenging than anticipated. Among the challenges encountered are:
- non-standard documentation of rights across the National Library's collections;
- the large amount of paper-based rights information;
- the complexity of copyright legislation; and
- resourcing.

Rights data

One of the early tasks undertaken was an examination of existing catalogue data to see what rights data was included in records. Most collection areas have not routinely included this information when items are catalogued, with the exception of our Oral History Branch. This meant there was limited rights data in accessible electronic format for migration to the new system. An enhanced awareness of the importance of including rights information in the catalogue record is emerging.

Although all special collection areas are within the Australian Collections and Reader Services Division, there has been significant variation between areas and over time in:
- the documentation of rights information;
- the documentation of permissions obtained by users;
- the conditions imposed on access and use by donors; and
- the application of risk management approaches in allowing access to and use of unpublished material.

The special collections areas are moving towards more standard approaches to these issues, a process which will be aided by the RMS.

It is anticipated that it will take many years for the RMS to provide real benefits to some areas. In particular, the manuscripts collection, with over two million individual documents, recognised that the lack of item level descriptions would be a problem.

However, they have been keen to be involved, recognising that in time an online RMS will allow significant amounts of collection management information to be readily accessible.

The inclusion of links into the Library's electronic records system, TRIM, has led to a large scanning load in an effort to put records into the system so documents about rights agreements and permissions could be linked to the RMS records. As the system becomes more widely used, and staff see the value in being able to easily link to these source documents, the demand for conversion of such documents will escalate.

Copyright

The copyright status calculation algorithm was released at an early stage in the first phase of the project. Not unexpectedly, because of the complexity of both copyright law and the materials in our collections, examples which force closer examination of the algorithm and catalogue records have been uncovered with great regularity. Some IT staff have become very familiar with the intricacies of the Copyright Act!

It is not enough to know that copyright for a recent book is publication year plus 70 years. Copyright can vary according to format, publication status, creator death dates, whether it was created by a government agency, age of the material and other factors. When the Free Trade Agreement was signed by Australia with the United States, the term of copyright was extended for a further twenty years, but only for material still within copyright. Rights can be held by one or more creators, the publisher, a literary agent, a descendant, or may have been transferred to the Library. For digital materials, there may be additional copyright issues in terms of format shifting and preservation copying.

Resourcing

On a more general note, the complexity of the project was underestimated initially, with the project manager being appointed in a part-time capacity. This created problems as the project progressed, with other high priority tasks limiting time spent on the project. The necessity for the system developers to work on other tasks for which they also have responsibility has also impacted timelines, as did staff turnover.

The future

In September 2007, the first release of the RMS was finalised and put into production. The next stage will focus on rightsholders, and is due for completion in the first quarter of 2008. Also during 2008, the final release of the current RMS project will see the implementation of a search function across all collection items.

Already it is apparent that workflows and practices within special collection areas are being modified to give more emphasis to standardisation and updating of rights information in catalogue records. Quantification of paper-based records has allowed collection managers to start planning for conversion of essential records, while the project team developed guidelines for efficient processing of this data.

Future National Library of Australia initiatives will enable public access to some elements of rights information to inform users about use of the collections, and facilitate more efficient document access, ordering and delivery.

References

California Digital Library/Library of Congress 2007. Data elements needed to ascertain copyright facts (MARC Discussion Paper No. 2007-DP05). Available from <http://www.loc.gov/marc/marbi/2007/2007-dp05.html>. [21 January 2008]

National Library of Australia 2007. IT Architecture Project Report. Available from <http://www.nla.gov.au/dsp/documents/itag.pdf > [21 January 2008]