**Citation**
Wells, D. 2009. Implementing Digitool 3 at Curtin University Library. In: ANZREG (Australia & New Zealand Regional ExLibris Group), 23rd Jan 2009, Sydney.

*Paper for ANZREG, 23 January 2009*

# IMPLEMENTING DIGITOOL 3 AT CURTIN UNIVERSITY LIBRARY

David Wells, Manager, Resources and Access, Curtin University Library
d.wells@curtin.edu.au

## Introduction

Curtin University Library is using Digitool 3 for Curtin's institutional repository, for our digital thesis collection, and also to manage several archival research collections.

This paper describes how we implemented these different products in DTL 3 with some background about how we set things up to achieve results we wanted in terms of system architecture and configuration.

I would like to the take opportunity to thank Natali Koifman of Ex Libris for her guidance and contribution to the project. I would also like to acknowledge the input of the project team, especially Jarvis Cochrane, Matthew Robinson, who did the programming for the repository conversion, and Colin Meikle, who has been largely responsible for the Resource Discovery configuration.

Curtin's implementation if DTL has been a lengthy process. The decision was made to use Digitool for Curtin's institutional repository and digital thesis collections during 2006 for implementation in the 2007 planning year. Actual implementation for the theses took place in April 2008 and for the repository in November 2008.

The archival research collections project was planned for the 2008 planning cycle. The start was delayed because of delays to implementation of the repository, and the project is still not complete. We are currently expecting to go live by the end of February 2009. Our experience with thesis and repository data was certainly valuable when we looked at archival collections and has kept further delays to part two of the project to a minimum.

## eSpace

Curtin's collection of repository items and digital theses is known collectively as eSpace. Material for eSpace was converted from two previously existing systems.

1) Digital theses: these were digitised versions of approximately 700 research theses (Master's and doctorates) submitted by students at Curtin. These were previously handled through the ETD-db software developed by Virginia Tech (http://scholar.lib.vt.edu/ETD-db/index.shtml).

2) A repository of approximately 2000 research papers by Curtin academic staff (eSpace). This was managed through the EPrints open source software

(http://www.eprints.org/). (This collection was previously known as eSpace and this name was continued as the name for the new combined collection.)

These two collections were merged into a single system in DTL3 and handled within a single admin unit (ERA02). The fields in the two legacy systems were mapped to Dublin Core and reconciled to each other. Basic fields – author, title, date, abstract -- are shared, but each also uses a lot of unique fields. As DC as such was inadequate to reflect the complexity of the data we also included a large number of locally defined dcterms tags.

Further complexity for the migration process with repository data came from the need to convert two separate data streams: 1) the existing live data from the EPrints database – this was migrated as a one off process, and 2) regular dataloads from the University's research management database (Script), which are brought into DTL through the ingest process and then need to be checked and enhanced by Library staff before items can be made publicly available. Objects are ingested with a status of 'archive', which is changed to 'view' at the appropriate moment. We maintain numerous fields in the metadata which are there for reporting purposes and to track workflows. As part of the conversion process we also had to write a number of reports for internal purposes based on these fields.

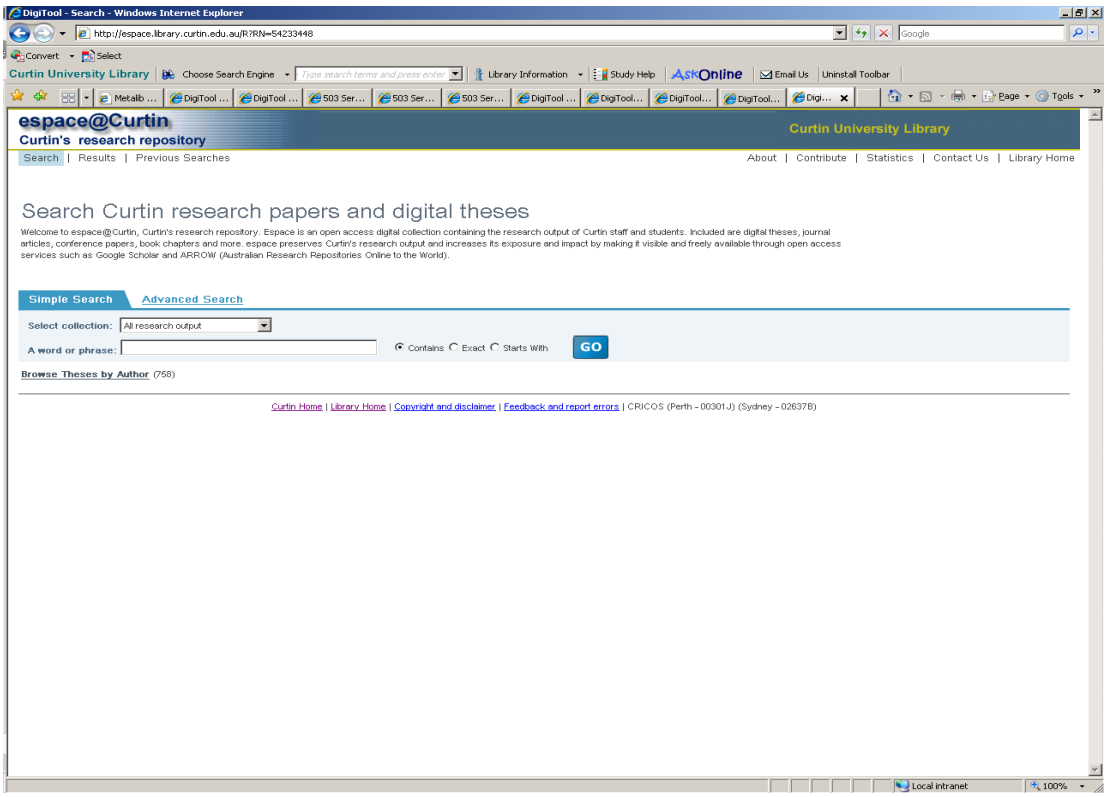


*Fig. 1. Entry point for new eSpace.*

The entry point for the new eSpace is a Basic Search box. The aim was to provide a clean screen with some basic information about what eSpace is. Some links along the top right of the screen point to further information, how to contribute, statistics, etc. Below the search box is a link to 'Browse Theses by Author'. This uses the DTL Collection Management functionality to retrieve all of the theses, sorted by Author, and was included in response to requests to include a browse list to replicate

functionality in the previous system. We may in the future extend the use of collections to other criteria – e.g. subject, university department or creation date, and include repository data. Collection Management cannot be used to create an alphabetical list of research articles, however, as these often have more than one author, and only the first author would be included.

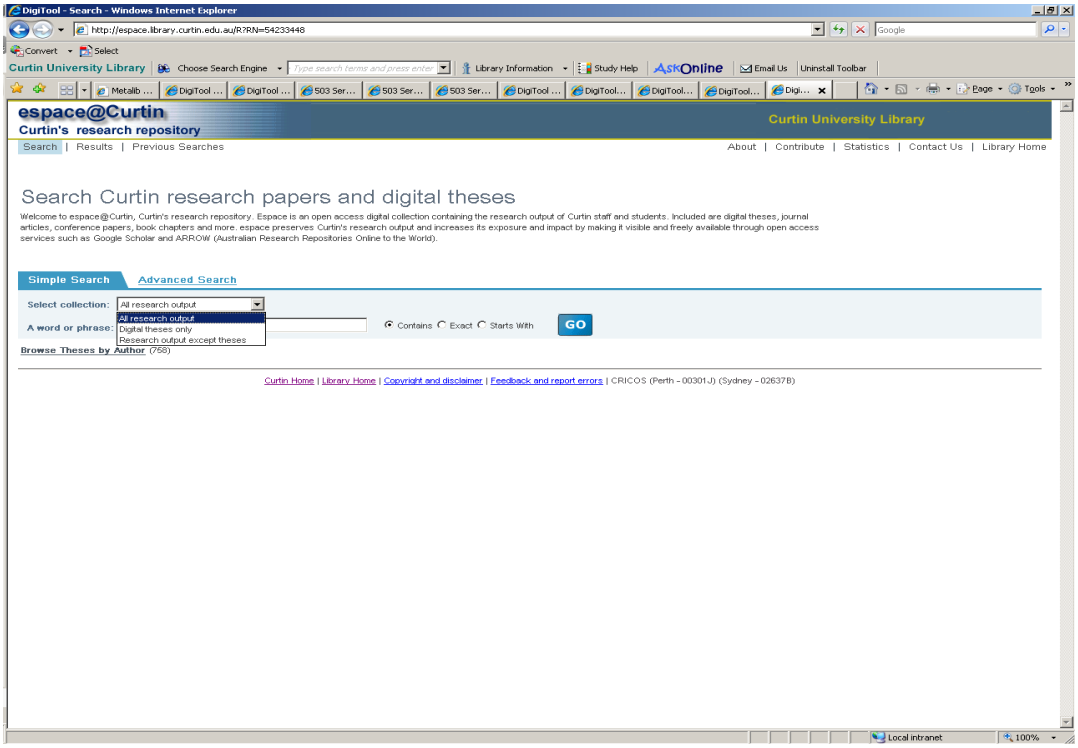


*Fig. 2. Search bases within eSpace*

Users can choose to search the whole of eSpace or just theses or just research papers. These are separately defined search bases within the ERA01 admin unit (Fig. 2).

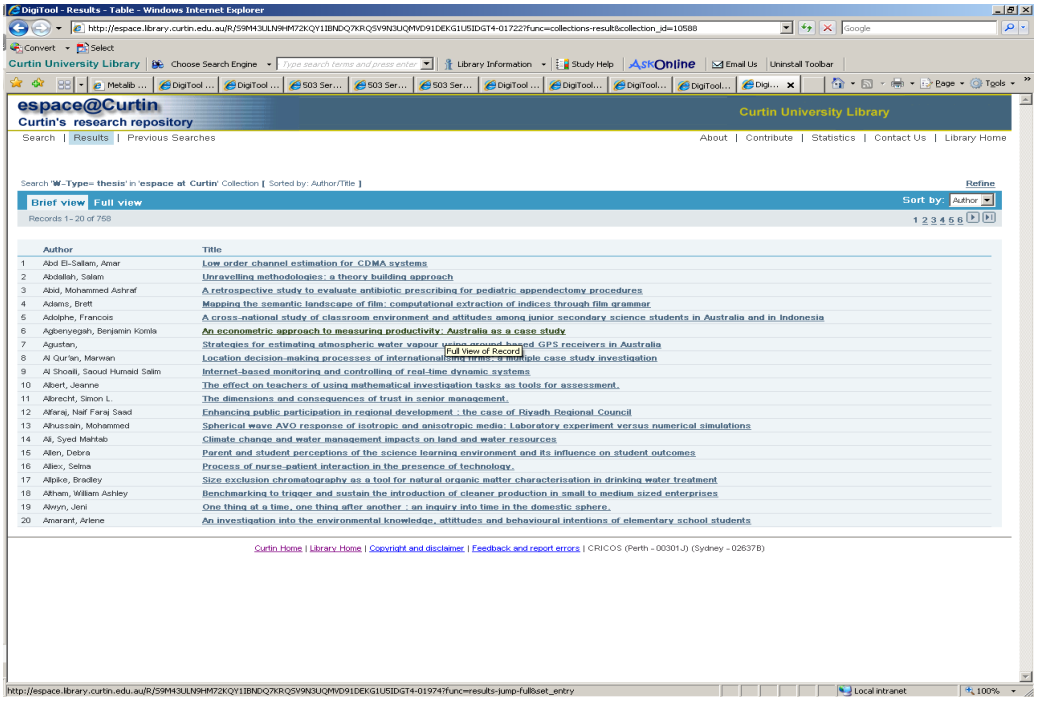


*Fig. 3. eSpace Brief Results screen*

Fig. 3 shows the Brief Results screen from the Browse Theses by Author link. We have kept the screen as simple as possible here by removing some links from the default version – e.g. links to digital objects, print and email options. We do not require users to log on, and have also removed the e-shelf.

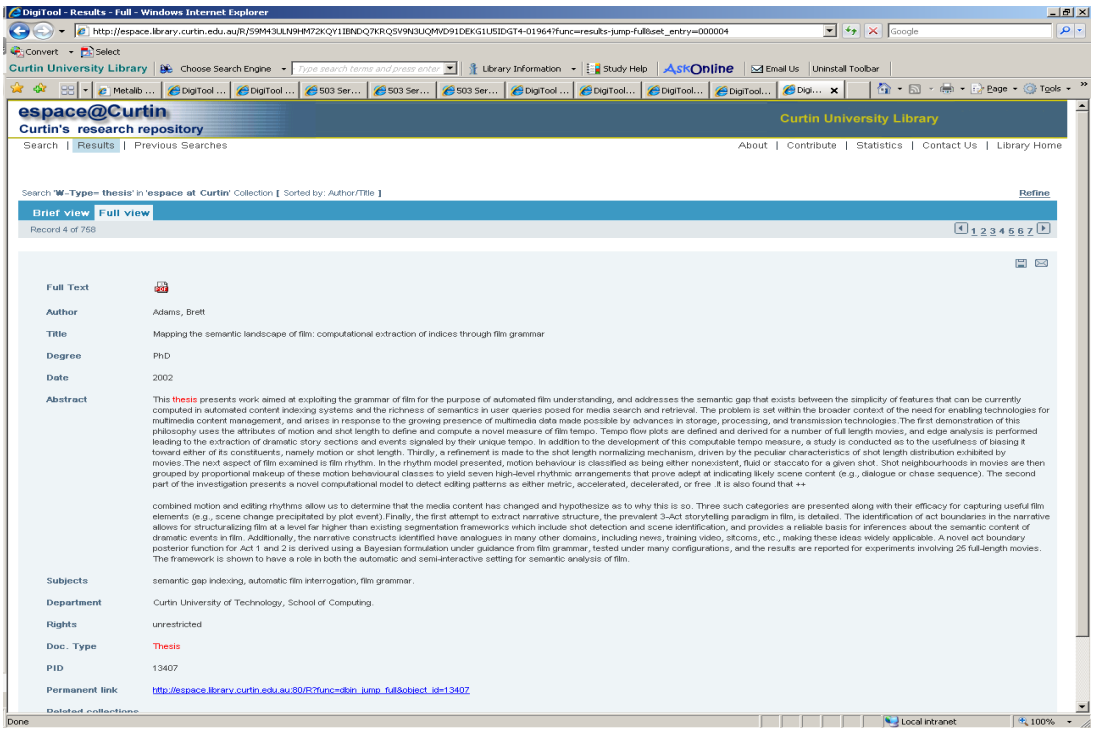


*Fig. 4. Full view of a thesis record*

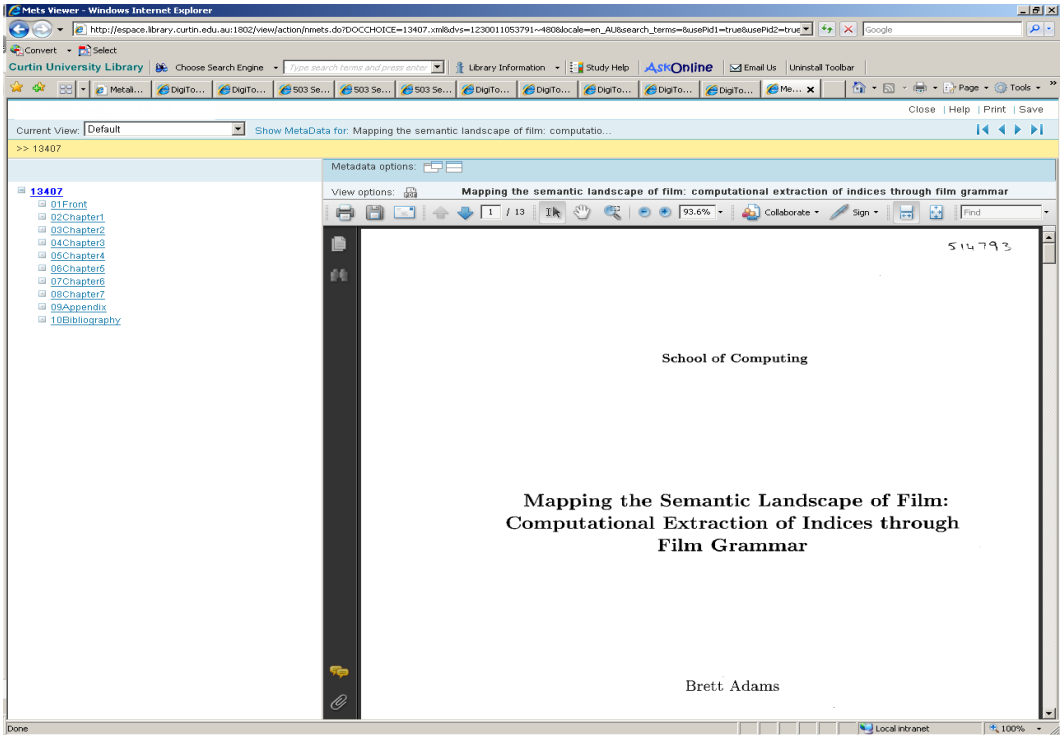


*Fig. 5. Object view of thesis record*

Theses were converted as complex objects in DTL with a separate pdf file for each chapter. We have subsequently decided to use single file for newly added theses unless they are particularly large.

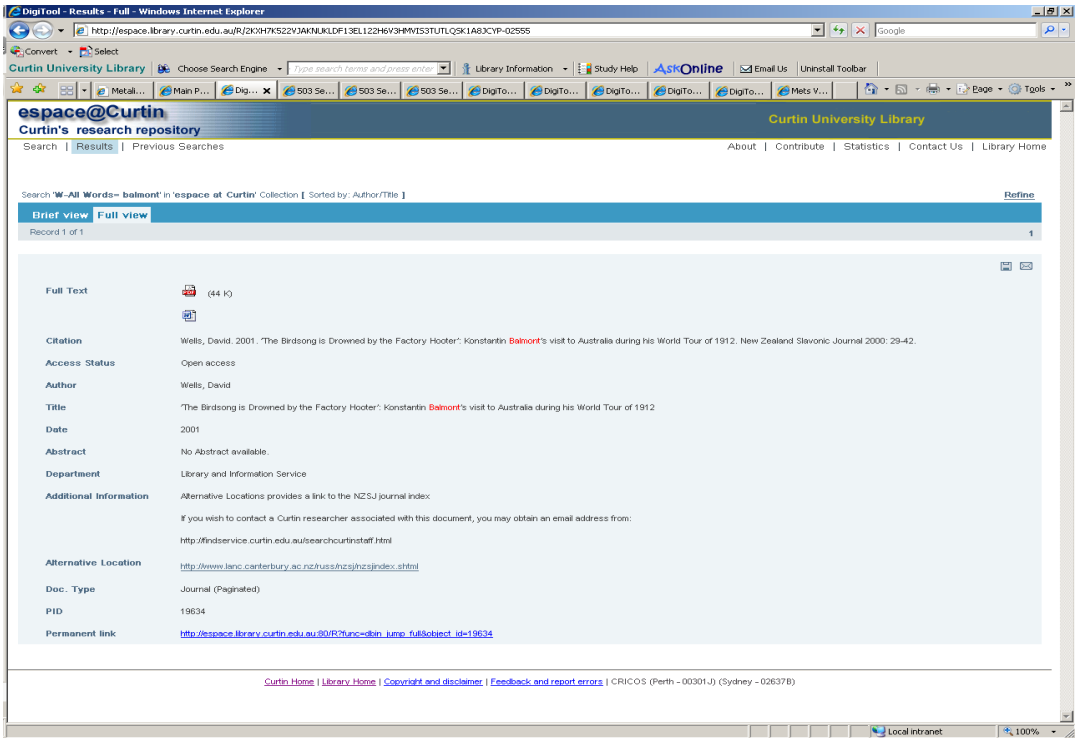


*Fig. 6. Full view of a research paper record from institutional repository*

The Citation field was generated automatically using relevant fields in the data during the conversion and continues to be created as part of the ongoing load from the

University research management database. Each record has an Access Status field indicating any restrictions on the text. Another field records information about alternative locations of the text when these are available. When appropriate, the institutional repository records make use of 'manifestations' functionality to bring together different versions of a text – for example author's version and published text of the same work. Access rights metadata may be applied to restrict specific documents. So far we have only used restriction by IP address, e.g. to Curtin IPs only or to processing staff only. In the record illustrated in Fig. 6 the Word document is available only to Curtin IPs. We are expecting that this functionality will also be used to satisfy the requirements of the Labor government's Excellence for Research in Australia (ERA) project if there is a need to link to documents for evaluation purposes.

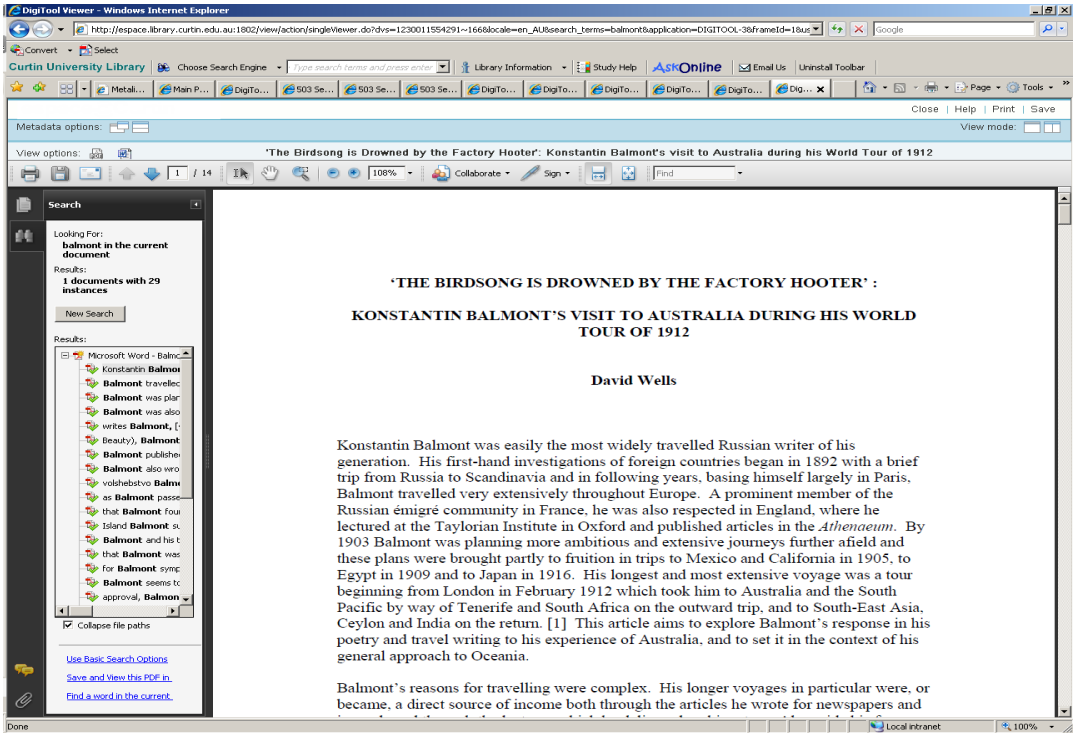

*Fig. 7. View of pdf version of repository document*

Note that we are using different viewers for thesis and repository items.

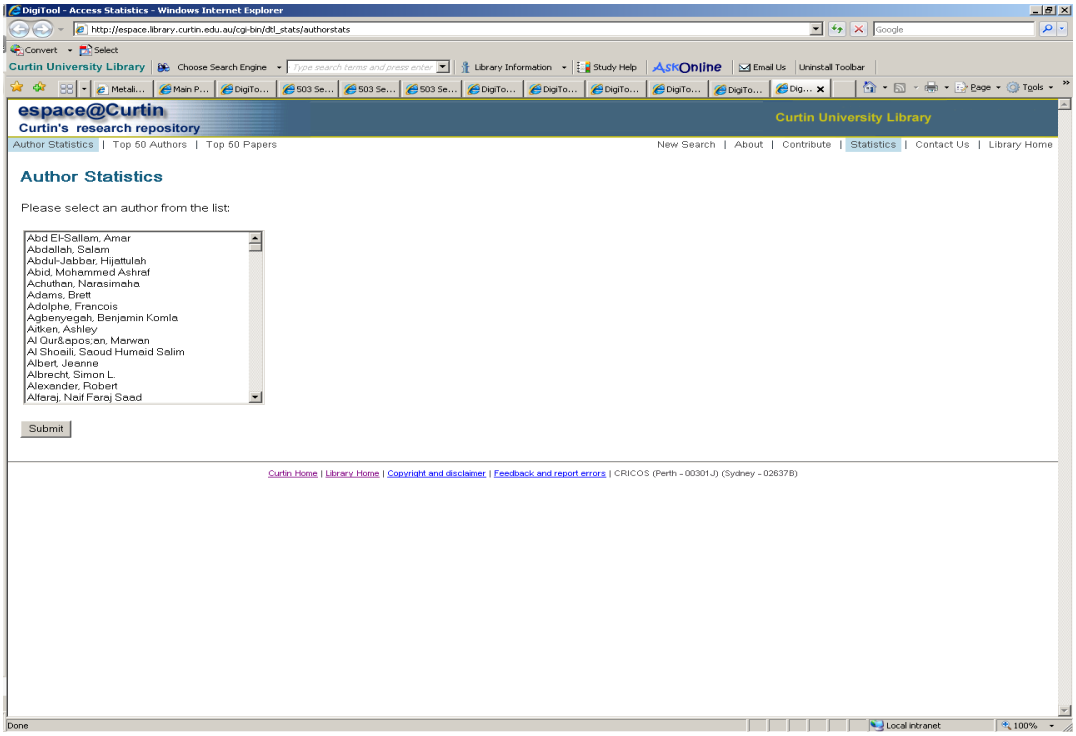


*Fig. 8. eSpace author statistics*

This is what you get if you click on the Statistics link from the eSpace search or results pages. Statistical information is particularly vital in the growing climate of research assessment and promotion in present-day Australian higher education. In order to capture statistics counts we have used in-house programming to embed a transparent image into the full record display, and added a javascript element to the link from the metadata to the digital object. This allows us to capture the PID relating to metadata or documents being viewed. The PIDs and count figures are logged in an external Oracle table and mapped back to authors and titles for display. The Oracle table also includes legacy statistics from our previous repository, allowing us to provide a historically continuous overview of viewing patterns of papers in our eSpace system. Fig. 8 shows an alphabetical list of Curtin authors.

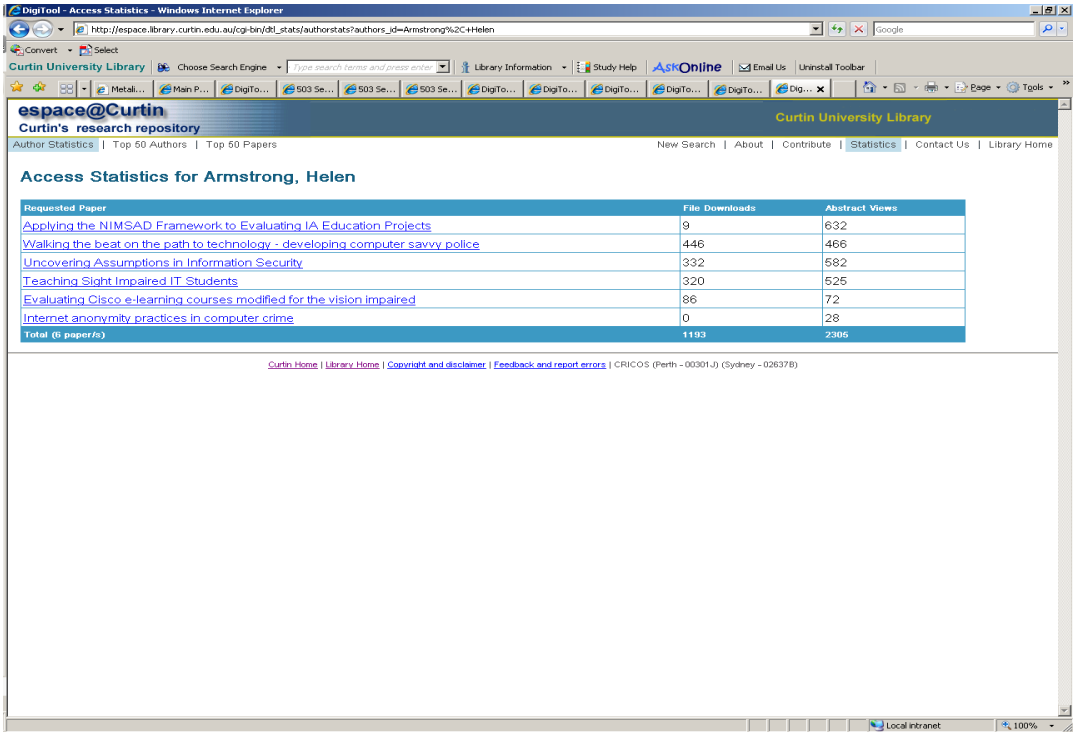


*Fig. 9. eSpace statistics for an individual author*

If you click on an author, you get a list of papers and counts of the number of times the full view of the catalogue record has been called up and the number of times the article has been downloaded (Fig. 9).

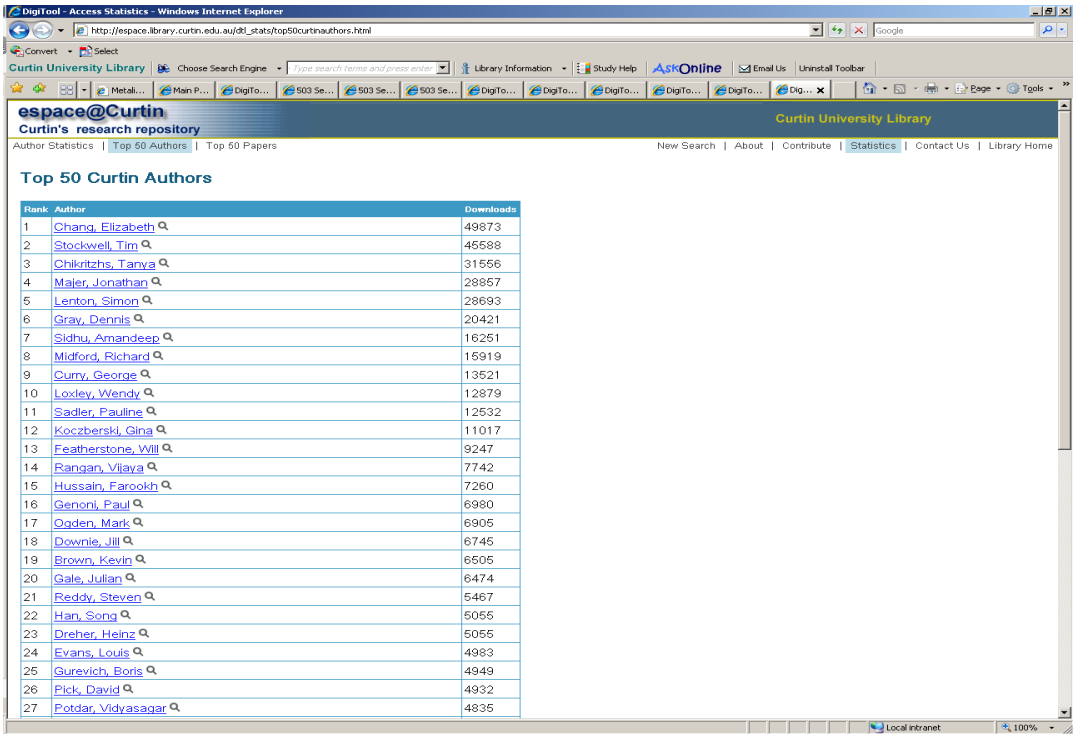


*Fig. 10. Top 50 Curtin authors*

Links form the statistics pages go to listings of downloads for the top 50 Curtin authors and the top 50 Curtin papers.

From here the magnifying glass link goes to the author statistics page for the author in question. Clicking on the author's name performs a search in Resource Discovery on the author in question.

**Curtin Archival Research Collections**

Phase two of Curtin's DTL 3 implementation project concerns the Curtin Archival Research Collections. The largest and oldest of these is the archival material of the John Curtin Prime Ministerial Library (JCPML). The JCPML is Australia's first Prime Ministerial Library, conceived on the model of the presidential libraries of the US and is broadly devoted to the life and career of Australia's wartime prime minister, and also more generally to Australian wartime and pre-war politics, and to the social and labour history of the period. The material in the JCPML archive is diverse in terms of medium, comprising contemporary documents and photographs, supplemented by audio material – recordings of Curtin's speeches, and oral histories collected from people who knew Curtin, for example; video; and a small collection of realia. Whenever possible this material has been digitised and made available to the public.

JCPML was one of the first sites to use Digitool for an archival collection and migrated from the previous separate archival management and public access systems to Digitool 2.2 in 2004. At present this material is still running in Digitool 2.4.

Building on the experience gained through establishing this collection, the staff of the JCPML have subsequently taken on responsibility for the development of a number of other archival initiatives and their management within the Digitool system.

- The Geoff Gallop Collection, which contains the personal papers of a former premier (2001-2006) of the State of Western Australia,
- The Carmen Lawrence Collection (premier of W.A. 1990-1993),
- The Elizabeth Jolley Research Collection, which is an online bibliography of works by and about the prominent West Australian novelist Elizabeth Jolley (1923-2007), with links wherever possible to the texts of relevant documents.

Further projects are planned, notably the West Australian Folklore Collection, and a collection of materials relating to the round-the-world voyage of the yachtsman Jon Sanders.

The archival collections will comprise a second admin unit within Digitool 3 (ERA01) and are encoded using MARC. The data has been converted by Ex Libris, incorporating as far as possible the functionality we had developed in DTL 2.

The Curtin archival collections in DTL have two distinctive features.

- As Digitool is used as a catalogue of the physical archives as well as a vehicle for making digital files available to the public, there are large numbers of metadata records which do not have digital files attached.

- The archival metadata records are linked hierarchically to reflect the logical structure of the archives. Following established standards for archival description, there are records for four different levels: collection, series, file and item, with digital objects attached only at the lowest level. In Digitool 2.4 the different hierarchical levels are linked using 774 tags placed in the 'parent' records. These create browsable links in the OPAC allowing simple navigation up and down the hierarchy of 'parents' and 'children'. The complexity of the different archives varies considerably. The John Curtin archive, for example, has approximately four hundred collection level records, whereas the Elizabeth Jolley Research Collection has only one.

As part of the conversion to Digitool 3, Ex Libris wrote programming to translate this hierarchical structure through a series of 'nodes' within Collection Management. At the same time, Ex Libris has also created two additional levels of hierarchy that will sit above the existing structure: first to distinguish between the different archives in our system, and second to break down the larger archives alphabetically, using a collection node to bring together records of creators whose names begin with each letter of the alphabet. This use of the hierarchical collections mechanism will allow us to keep the archival structure firmly in the view of our clients.

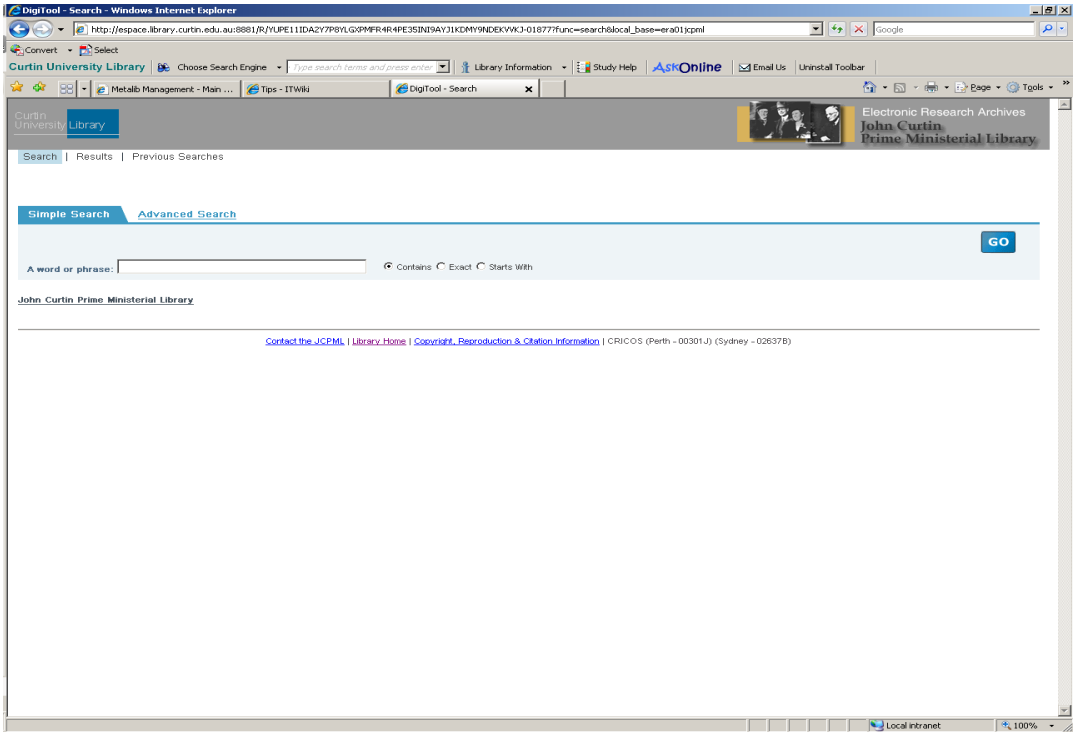


*Fig. 11. Entry point for John Curtin Prime Ministerial Library*

The pages shown here for the Curtin research archives are draft pages only and may change slightly before we go live. Although they are all in the same admin unit the different archival collections have been defined as separate search bases, and have each been given their own set of web pages with separate branding. Fig. 11. is the start page for the JCPML collection. As with eSpace we have started from a Basic Search box with minimal clutter. This is restricted to the JCPML search base, and contains a top level link to the collection hierarchy for JCPML material. This was

created by hard coding the collection id into the web page. Otherwise, as the collection management functionality itself can only be configured at the level of admin unit, not at search base level, we would not be able to exclude collections not specific to the particular search base we are in.

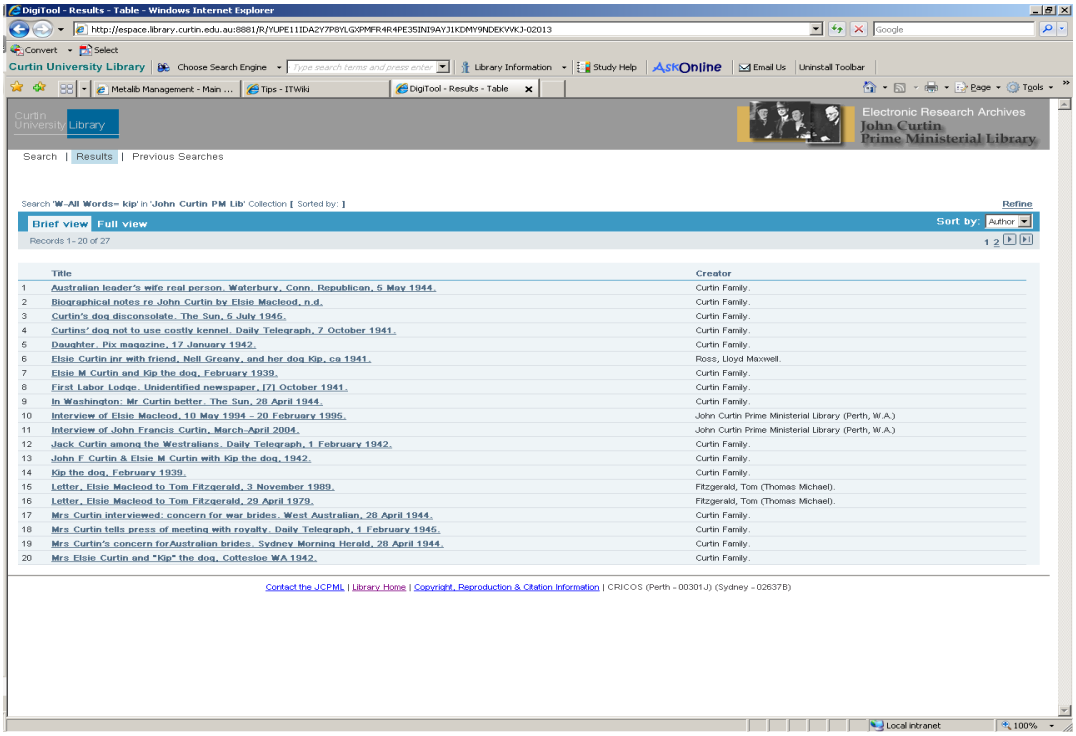


*Fig. 12. Research archives brief view*

Fig.12 shows the Brief view display after searching for 'Kip' – the name of John Curtin's dog. Note that the order of fields is different from the equivalent eSpace display, with title first, instead of author. Creator here is the name of the archival collection to which the document belongs, not necessarily the person responsible for creating the document itself.

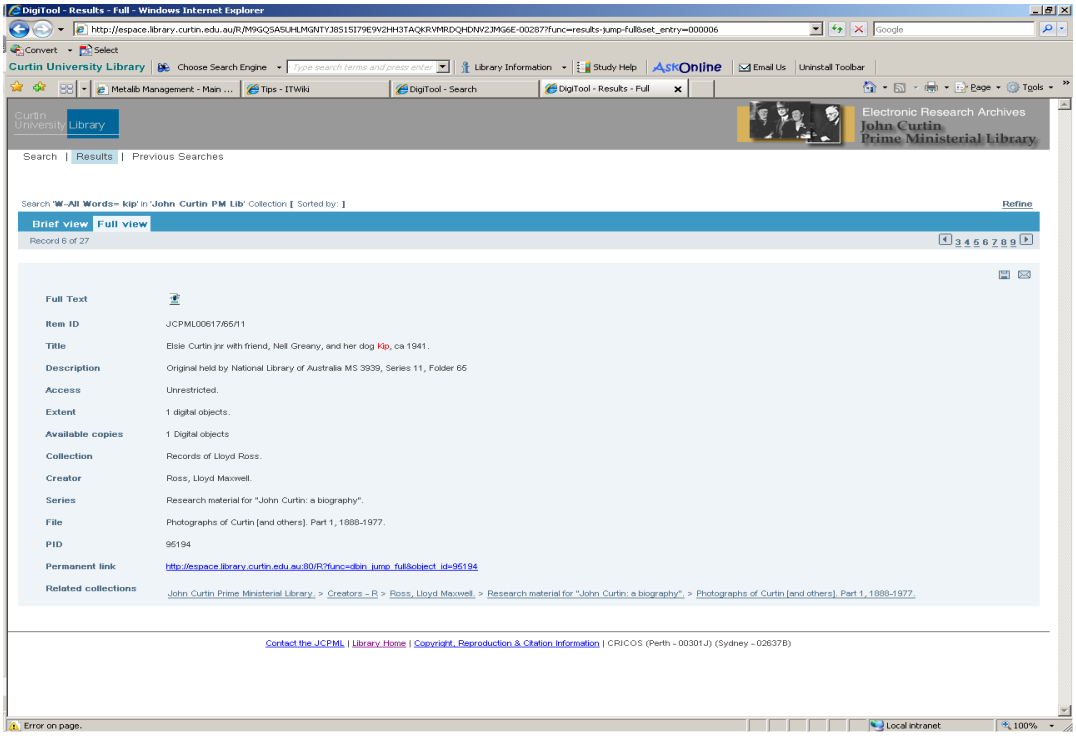


*Fig. 13. Research archives full view*

Fig. 13 shows the Full view display. Note that the fields displayed are substantially different from those in eSpace records: for example there are archival tags relating to extent and availability. Note also the Related Collections link at the bottom of the screen – a breadcrumb trail relating to the collection hierarchy.


*Fig. 14. Research archives object viewer*

The object viewer used with the Curtin research archives is the one developed at Curtin for use with DTL 2 (Fig. 14). This was retained partly in order to allow separate branding for individual archival collections.

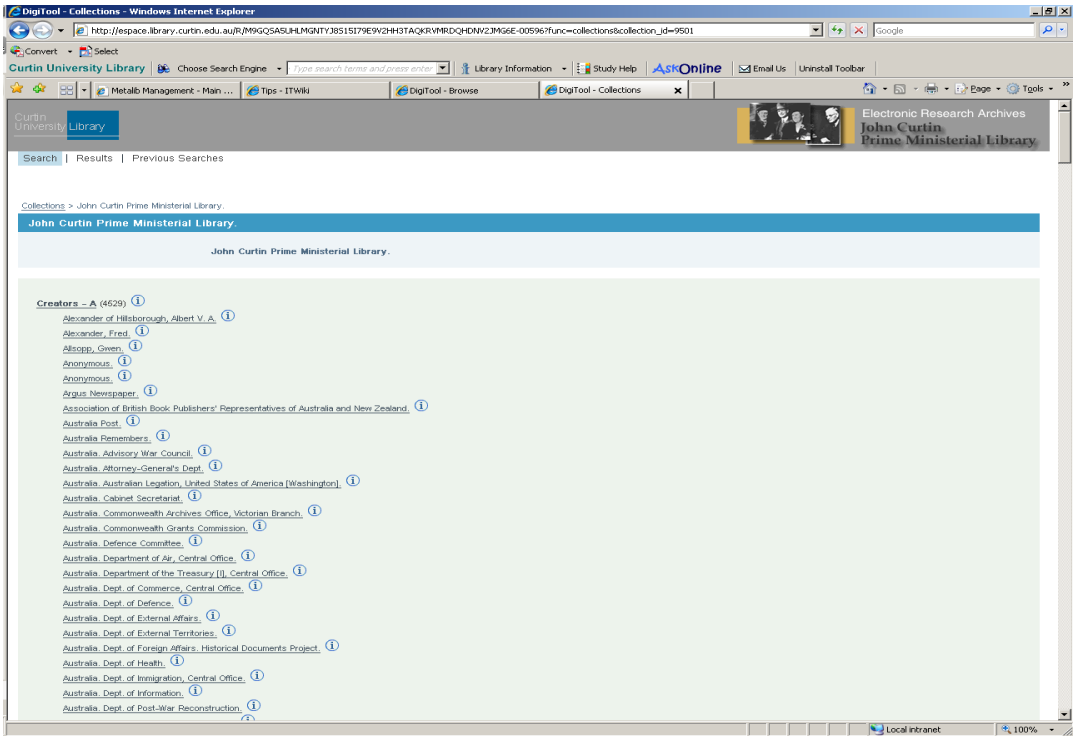


*Fig. 15. Collection hierarchy*

Fig. 15. shows the result of clicking on the top level collections link from the basic search page (Fig. 11). The creators of the different collections are subdivided alphabetically and presented in alphabetical order.

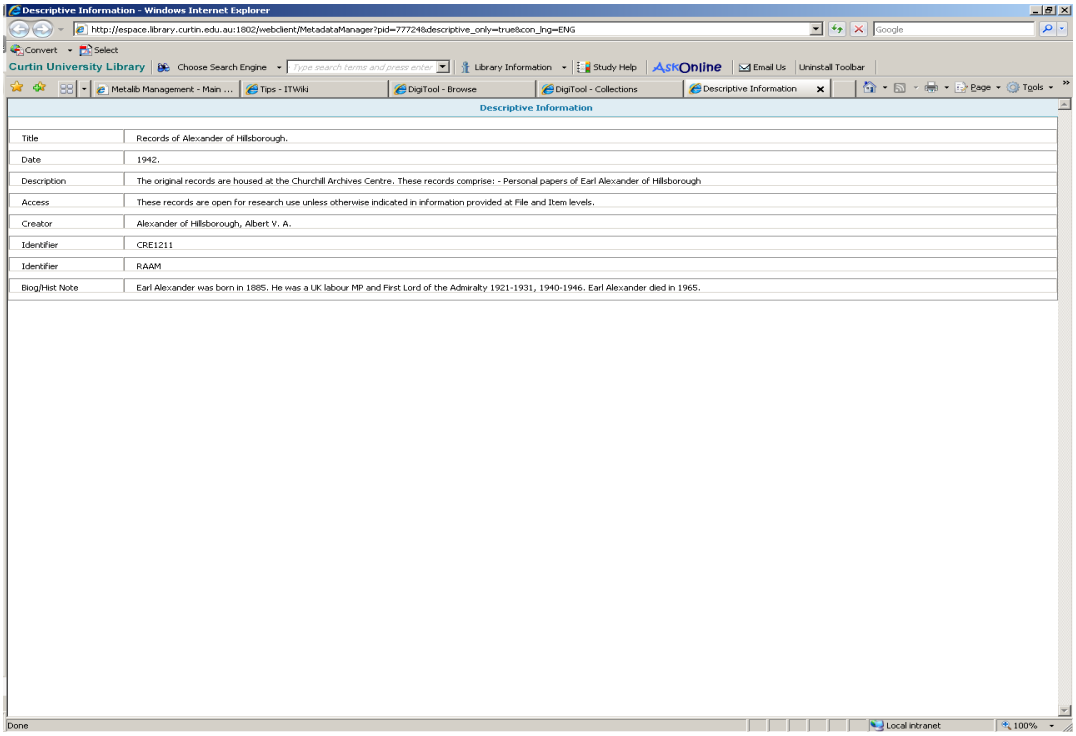


*Fig. 16. MARC record view from collection hierarchy*

The i button next to each collection links to a customised version of the MARC record (Fig. 16).

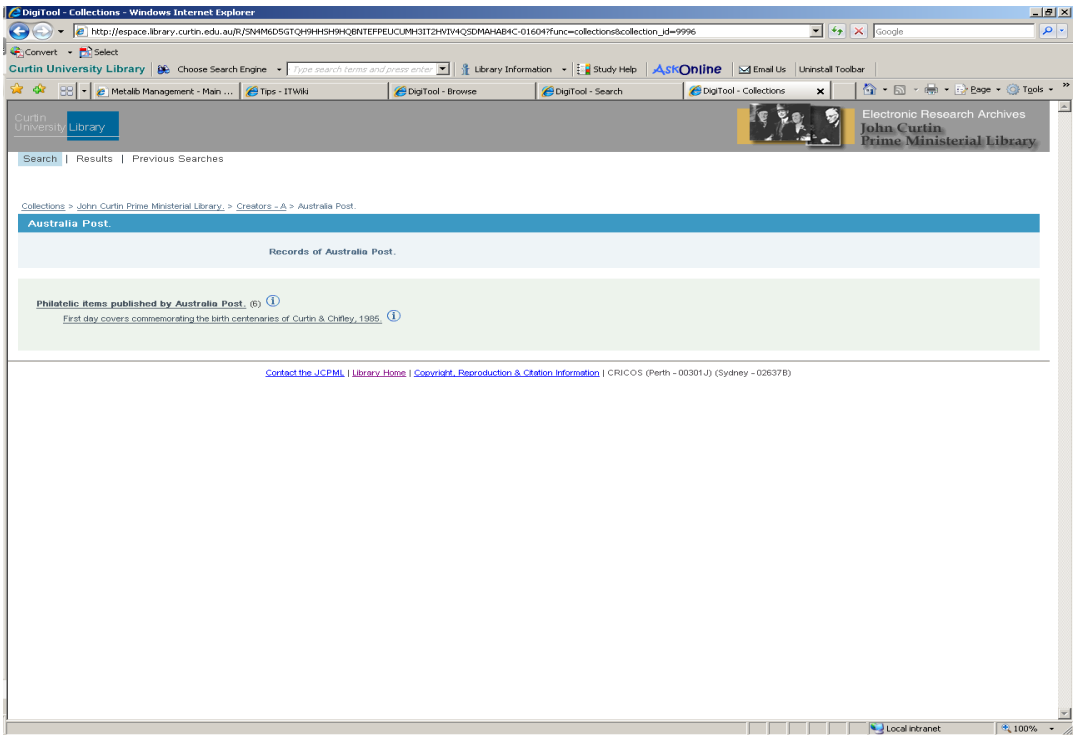


*Fig. 17. Hierarchy view*

Clicking on a creator name takes you to a hierarchical view including the whole structure – a view looking downwards in the body of the screen and a breadcrumb view upwards towards the top. Fig 17 shows this view for the records of Australia

Post. Clicking on the lowest level of the hierarchy takes you to the catalogue records associated with the relevant objects.

Each of the archival collections has its own distinctive branding. Figs. 18-20 show the entry pages for Geoff Gallop, Carmen Lawrence and Elizabeth Jolley. Further pages follow essentially the same structure as for the JCPML material.

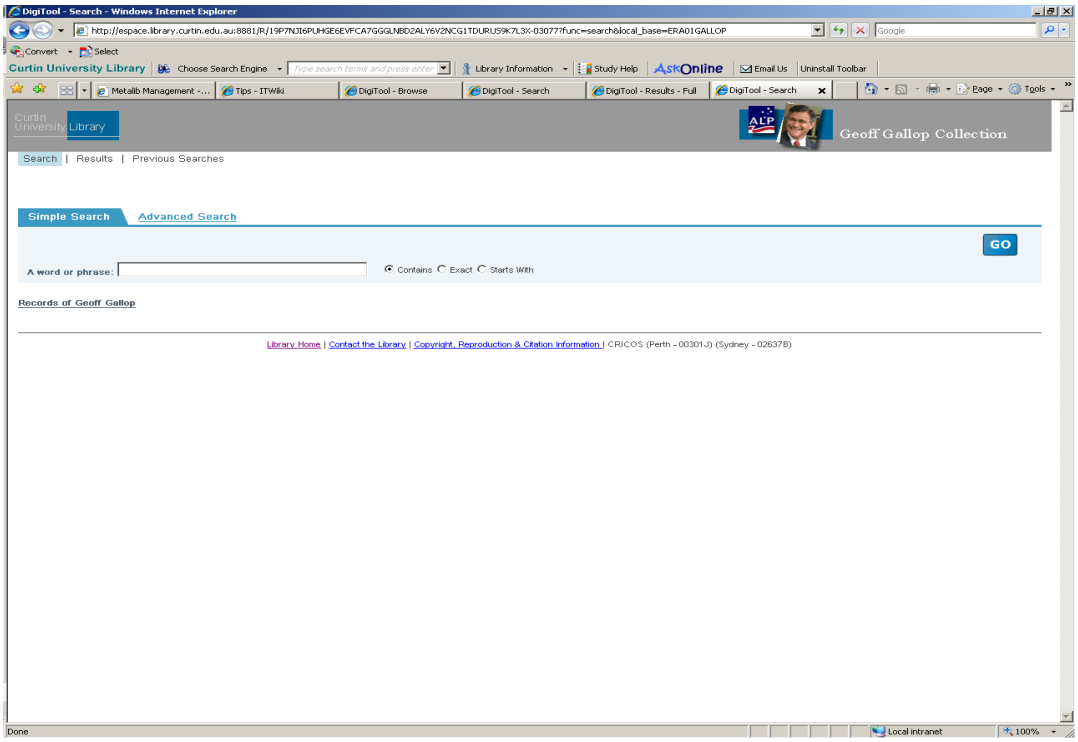


*Fig. 18. Entry page for Geoff Gallop Collection*

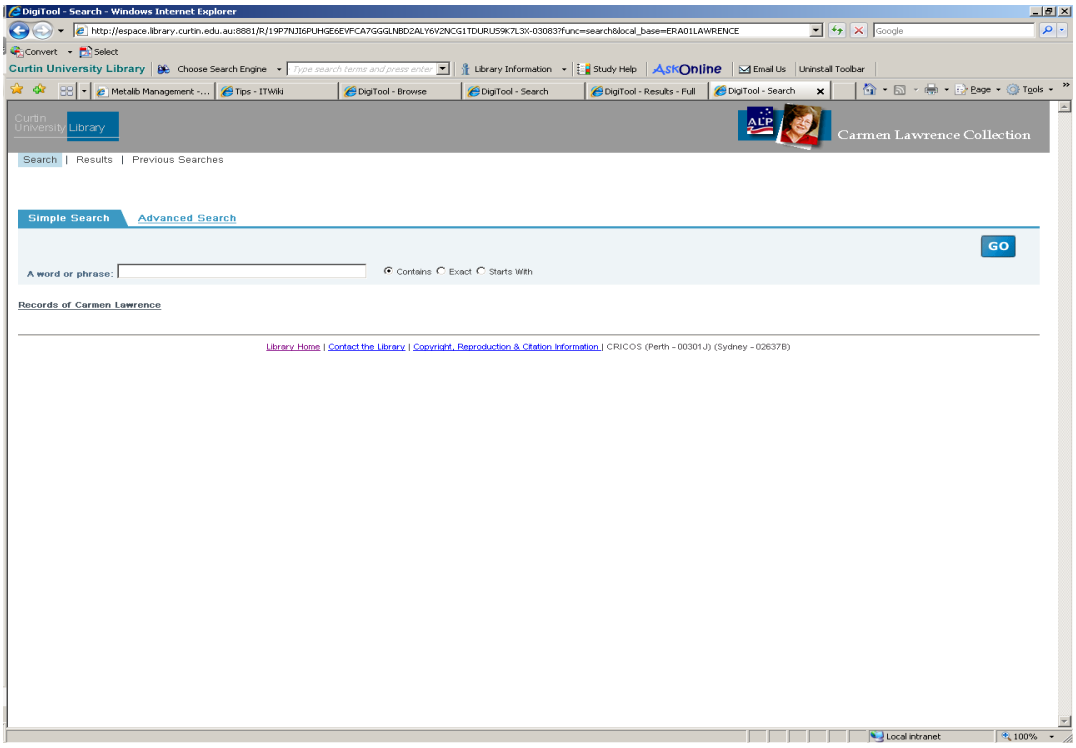


*Fig. 19. Entry page for Carmen Lawrence Collection*

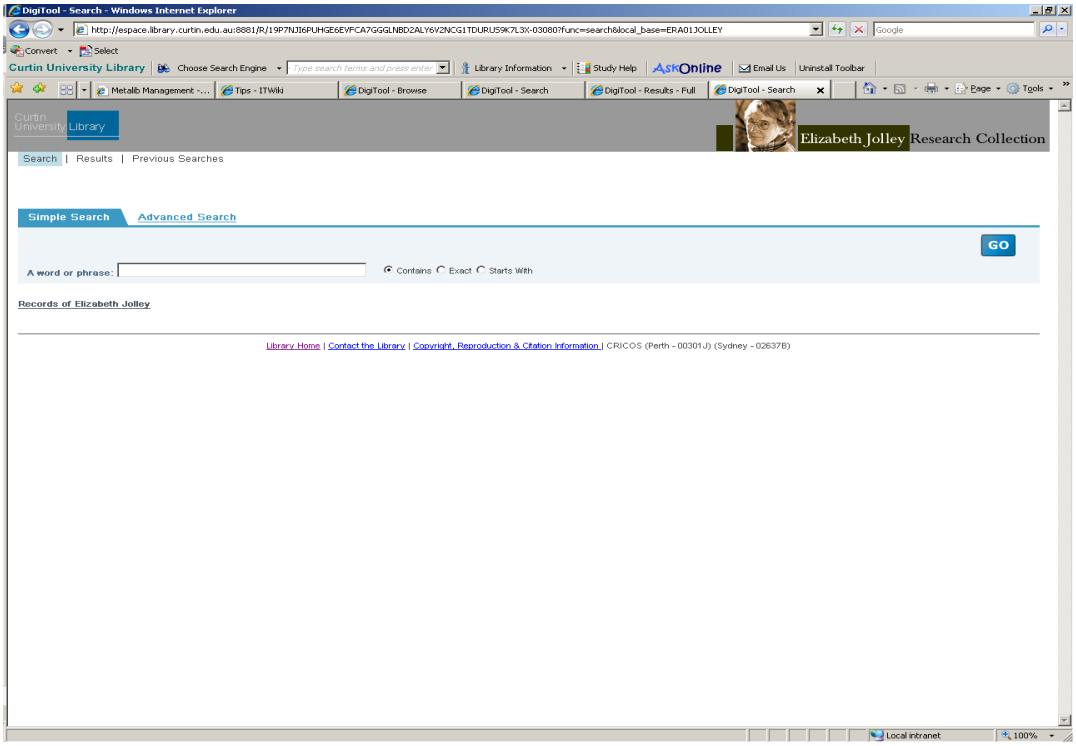


*Fig. 20. Entry page for Elizabeth Jolley Research Collection*

The URL for eSpace is http://espace.library.curtin.edu.au.

The URLs for the Curtin Research Archive collections will be made available shortly.