# Assessing the collective wealth of Australian research libraries: measuring overlap using *WorldCat Collection Analysis*

PAUL GENONI AND JANETTE WRIGHT

## Abstract

This paper reports the results of recent research examining the holdings of Australian research library collections recorded in the WorldCat database using OCLC *WorldCat Collection Analysis* software. The objectives of the research are:
1. To better understand the distribution of printed monographs amongst Australian research collections in order to assess the potential for enhanced collaboration in aspects of collection management.
2. To test the OCLC *WorldCat Collection Analysis* software in order to ascertain its value in comparing collection data based on the Australian research libraries subset of the WorldCat database.

The collections compared are the National Library of Australia; University of Melbourne; Monash University, and CAVAL Archival and Research Materials Centre. The data record the extent of overlap between collections, and the prevalence and distribution of single copies. The paper reflects on the use of *WorldCat Collection Analysis* software as a means of supporting the future management of Australian research collections. The research was undertaken as a pilot for a larger study.

IMPLICATIONS FOR BEST PRACTICE
- The results support those of previous overlap studies that point to the potential for significant de-duplication of Australian research collections
- *WorldCat Collection Analysis* software has the potential to enable far more detailed comparison of Australian collections than has previously been possible.
- Libraries, or groups of libraries, undertaking detailed collection overlap analysis will have available data that can be used to support collaborative collection management.
- Identification of last copies in the national collection will provide added confidence in their retention.

## Introduction

As library collections continue to transition from physical to digital formats, library managers are faced with the challenge of deciding on the medium-term and long-term storage of their print collections. For while libraries continue to acquire many new print items, there is a widespread acceptance that the proportion of new material acquired in this form will continue to decline. In addition to the increasing amount of content that is acquired in a digital form, research libraries and communities are also adjusting to the impact of mass-digitisation programs that will potentially absorb at least some of the demand for access to print materials.

As a result research libraries are under increasing pressure to manage their "legacy collections" of print material in the most space and cost-effective manner possible, while continuing to provide efficient access to items from these collections. This issue is particularly acute in research libraries where there is a responsibility to ensure that the formidable print-based collections remain secure and accessible, while at the same time freeing-up space for new technology dependent uses (Sharp 2009).

In this environment it is imperative for research library managers to have reliable data on which to base decisions relating to the storage, disposal and digitisation of items in print collections. Increasingly these decisions are being taken from a basis of collaboration, with a view to meeting the needs of a group or network of libraries, while reducing the burden on individual libraries within those systems. This reliance on collaboration means that data used to underpin decision making should also be system-wide in order to provide the most relevant evidence to support those decisions.

The Australian research library community is adequately cohesive, and sufficiently supported by an existing collaborative discovery and delivery infrastructure, to enable it to potentially function as a single network. In this circumstance there is significant benefit in approaching key collection management decisions on a whole-of-system basis. This is particularly true of decisions relating to the sustainable management of low-use print materials—the so called "long tail"—where collaboration provides great potential for substantial savings in the cost required by long-term print storage.

This research is therefore intended as a pilot study for a wider analysis of the incidence of overlap and last copies within Australian print collections. It is undertaken as part of an ongoing project investigating the long-term storage, discovery and delivery of legacy print collections for the mutual benefit of Australian research libraries and communities. It also forms part of a considerable body of international research and commentary looking at the possibilities for the transformation of print storage through increased collaboration (see for example: Vattulainen 2004; Gherman 2007; Payne 2007; O'Connor and Jilovsky 2008; Yoon and Oh 2008; Genoni and Varga 2009; Sharp 2009).

**Australian collection overlap studies**

Collection overlap studies are a standard method of investigating the relationship between collections and the distribution of items between two or more libraries. The data derived from overlap studies is useful for both the individual libraries included in the studies, and for the group of libraries whose collections are examined. Data relating to overlap and unique holdings can assist in making decisions related to the management of those collections. The types of decisions that might be influenced include collection development; last-copy retention; inter-library loan/ document delivery; disposal, and storage. Because overlap studies involve two or more libraries the data is particularly useful for libraries seeking to develop cooperative policies or processes relating to the management of their collections.

The capacity to conduct overlap studies depends on the ease and accuracy with which holdings can be compared. In recent years the development and implementation of increasingly large-scale, inclusive union catalogues has provided additional impetus for overlap studies by enabling them to be increasingly broadly-based and effective. The use of such catalogues does, however, raise issues relating to methodology and the completeness and accuracy of catalogue data (Rochester 1987), and studies based on national union catalogues inevitably encounter problems associated with inaccurate and incomplete data.

In Australia recent overlap studies have relied upon the holdings recorded in the National Bibliographic Database (NBD). These include a study conducted by the National Library in 2002 on behalf of the Higher Education Information Infrastructure Advisory Committee (Missingham and Walls 2003). This study investigated the overlaps between academic libraries on a state-by-state basis, and included both serial and non-serial holdings. Missingham and Walls encountered some of the frustrations of relying upon the NBD for examining overlap, noting that incomplete holding and duplicate records had the effect of "limiting the accuracy of any study based on a large collaborative catalogue" (p. 249).

A second major study focusing on academic library holdings was undertaken during 2002 and 2003 when the Australian Research Libraries Collection Analysis Project (ARLCAP) examined the overlaps in the South Asian and Indian Ocean collections of the Group of Eight libraries (serving Australia's most research intensive universities) and the National Library (ARLCAP 2004). The ARLCAP research is relevant to the current study in that it used the Automated Collection Analysis Services (ACAS) of OCLC. The ACAS undertook the analysis on behalf of ARLCAP relying upon the holdings data recorded in the NBD. The classification numbers of items were mapped to the WLN / OCLC conspectus in a manner similar to that used for the current research. The results were compromised to some extent by the low number of holdings (as low as 55% for one library) that had at that time been added to the NBD by participating libraries. Nevertheless the ARLCAP *Report* concluded that:

The project has demonstrated that the use of OCLC's ACAS to perform an automated collection analysis across several libraries is an effective tool for gathering data and structuring it according to WLN Conspectus divisions. The results provide a solid basis for further comparative analysis of the holdings, trends, and gaps in library collections.

Shortly after the completion of the ARLCAP research OCLC withdrew the ACAS service and announced that future collection analysis services would be based exclusively on the use of the holdings recorded in WorldCat and using the *WorldCat Collection Analysis* software.

OCLC *WorldCat Collection Analysis* software divides subject content according to the OCLC Conspectus. OCLC describe the Conspectus as "a framework to systematically inventory and describe library collections" (OCLC). The structure of the Conspectus is hierarchical, and is comprised of divisions (the broadest category), categories and descriptors. The divisions, categories and descriptors can be mapped to Dewey Decimal, Library of Congress, and National Library of Medicine classification schemes. Dewey Decimal mapping was used in this research as all four collections use DDC. There are 32 divisions within the OCLC Conspectus, and overlap data for 24 of these divisions was collected in the course of this study.

The most recent Australian overlap study was conducted in 2007 as part of the ongoing research project that is reported in this paper (Genoni and Varga 2009). The study examined overlap within the membership of CAVAL Ltd, a consortium owned by twelve Australian university libraries. The study also included the CAVAL Archival and Research Materials (CARM) Centre, a print repository providing a storage facility for member libraries and a document delivery service for the wider research community. The study relied upon an analysis conducted by the National Library of NBD holdings data of the relevant collections, and was limited to monographs in the Dewey Decimal range of 600-699.

OCLC WorldCat

OCLC WorldCat has become established as the foremost international union catalogue. As at September 2010 OCLC claimed the database consisted of over 203 million bibliographic records with 1.64 billion holdings provided by over 72,000 libraries (http://www.oclc.org/worldcat/statistics/default.htm). Given the amount of catalogue data that is federated in WorldCat it is not surprising that librarians and researchers have investigated ways in which this extraordinarily rich source can be used to support research investigating the nature of collections and to make decisions related to their management. The potential uses cover a wide range of library operations including collection management, with Lavoie, Dempsey and Connaway (2006) arguing that with the assistance of WorldCat,

> . . . data mining across library collections could open the door to new opportunities for shared collection management. Studies of holdings

patterns for institutional clusters at the consortium, regional, or even national level could reveal opportunities to reduce cross-collection redundancies and free up resources to fill gaps in collections.

Some of the reported research-based uses of WorldCat data include: identifying the distribution and characteristics of last copies to provide data for decisions relating to de-accessioning and storage (Connaway, O'Neill and Prabha 2006); making inferences about the level of audiences for which texts are intended (O'Neill, Connaway and Dickey 2008); assisting with collection development by testing the effectiveness of an approval plan (McClure 2009); and conducting a collection evaluation test by comparing strengths and weaknesses of different collections (White 2008). The *WorldCat Collection Analysis* software has been used to conduct "brief tests" of collection strengths and weaknesses (Beals and Gilmour 2006), and to support decisions relating to the withdrawal of material from storage facilities (Ward and Aagard 2008).

From July 2007 the National Library of Australia entered into an agreement with OCLC that covered all Libraries Australia subscribers. Under the terms of the agreement records in the NBD that have attached holdings are uploaded to WorldCat (with a small number of exceptions for records obtained from some commercial suppliers); and WorldCat records with Australian holdings are in turn uploaded to the NBD. The agreement did not give Australian libraries access to additional services such as *WorldCat Collection Analysis,* although the National Library noted that the arrangement with regard to exchange of catalogue data would "allow Australian libraries to benefit from OCLC research and development" (National Library of Australia).

**Research Methodology**

Aim

The aim of the current research design is on recording the extent of overlap between collections, and identifying the likely prevalence and distribution of single (last) copies in the collections of Australian research libraries. The particular objectives of the pilot phase are:

1. To better understand the distribution of printed monographs amongst Australian research collections in order to assess the potential for enhanced collaboration in aspects of collection management. This includes the use of high-end technologies to support seamless discovery and delivery for the purpose of interlibrary loan and document delivery.

2. To test the OCLC *WorldCat Collection Analysis* software in order to ascertain its value for comparing collection data based on the Australian research libraries subset of the WorldCat database.

Methodology

*WorldCat Collection Analysis* software was used to undertake a study of holdings of single (last) copies in, and collection overlap between, a subset of Australian research library collections. 'Australian research libraries' in this context was defined as the members of Council of Australian University Librarians (CAUL); the Australian members of National and State Libraries Australasia (NSLA); and the CARM Centre. The collections included in the study were the libraries of The University of Melbourne (UM) and Monash University (Mon) representing CAUL; the National Library of Australia (NLA) representing NSLA, and the CARM Centre.

The data mined from WorldCat were intended to identify:
- The number of unique titles held by each library. Four results are possible:
    UM; Mon; NLA; CARM
- The number of titles held by any two of the libraries. Six results are possible:
    NLA+UM; NLA+Mon; NLA+CARM; UM+Mon; UM+CARM; Mon+CARM
- The number of titles held by any three of the libraries. Four results are possible:
    NLA+UM+Mon; NLA+UM+CARM; NLA+Mon+CARM; UM+Mon+CARM
- The number of titles held by all four libraries. One result is possible:
    NLA+UM+Mon+CARM

As noted, the quality of data in union catalogues has been a problem with many overlap studies, and using WorldCat does not avoid these problems (Orcutt and Powell 2006). Holdings data in WorldCat may be incomplete (for example not all records have been uploaded); inaccurate in a fashion which prevents matching of the same item resulting in duplicate records; or contributing libraries might have different cataloguing practices (e.g. with series titles) that prevent similar items from being identified. It is, for example, estimated that some 50,000 to 70,000 records for CARM Centre holdings that are recorded in the Libraries Australia database have not been able to be uploaded to OCLC due to system problems. There are indications that this is also true for the holdings of the university libraries included in this research. This will result in distortions to the overlap data and a likely understatement of the degree of overlap. The rate of duplication within this network of libraries will also be understated as this methodology does not count duplication within a collection. That is, multiple holdings of the same title by a single library will not be identified.

**Results**

The results presented in Table 1 were obtained by compiling the data from the 24 Conspectus divisions, plus those designated by the WorldCat collection analysis process as "unknown" (ie items for which a subject division could not be determined).

The Table presents data for the number of items that are held uniquely by each of the four collections, plus the extent of overlap as measured by items that are held by two, three, or all four of the collections.

Uniqueness and overlap

|  | Unique | % | Held by 2 | % | Held by 3 | % | Held by 4 | % | *Total* |
|---|---|---|---|---|---|---|---|---|---|
| CARM | 114,119 | 57.8 | 41,964 | 21.3 | 28,768 | 14.6 | 12,522 | 6.3 | 197,373 |
| UM | 617,006 | 47.8 | 437,636 | 33.9 | 222,514 | 17.3 | 12,522 | 0.9 | 1,289,678 |
| Mon | 458,421 | 41.7 | 397,888 | 36.2 | 231,275 | 21.0 | 12,522 | 1.1 | 1,100,106 |
| NLA | 1,594,816 | 70.5 | 420,678 | 18.6 | 232,826 | 10.3 | 12,522 | 0.6 | 2,260,842 |
| ***Totals*** | | | | | | | | | |
| *Holdings* | 2,784,362 | 57.4 | 1,298,166 | 26.7 | 715,383 | 14.8 | 50,088 | 1.0 | 4,847,999 |
| *Items* | 2,784,362 | 75.6 | 649,083 | 17.6 | 238,461 | 6.5 | 12,522 | 0.3 | 3,684,428 |

Table 1:  Unique holdings and overlap

The 3,684,428 items have a total of 4,847,999 holdings, with an average of 1.32 holdings per item. This indicates that there are some 1,163,571 duplicate holdings within the 24 subject divisions of these collections.

While this can be construed as a significant level of overlap, it is also noticeable that the level of unique items could also be assessed as being high, with 75.6% of all items having one holding only. The comparatively high level of unique holdings within the National Library has been noted in previous overlap studies that have compared the National Library with academic libraries (ARLCAP). This can be explained by the National Library's historical–but now reduced–role of collecting in depth for some international materials; and their continued commitment to the comprehensive collecting of Australiana, irrespective of the 'level' of the intended readership. In both cases this will result in the acquisition of material that is unlikely to be of interest to curriculum driven academic library collections.

The considerably higher rate of duplication within the CARM Centre collection (ie the high rate of holdings of items that are held within each of the other three collections) is likely explained by the presence of duplicate copies within the collections of member libraries, with de-duplicated copies being deposited with CARM. It might be assumed that these are likely to be textbooks or similar curriculum focused items.

By Conspectus division

Within the scope of this paper the results for three subject divisions are reported as examples of the type of data that can be readily extracted using *WorldCat Collection Analysis.* The subject divisions are Art and Architecture (211,880 total holdings); Sociology (238,461); and Medicine (250,041). These divisions were selected to represent the three broad disciplinary groupings of humanities, social science and

science; and because the number of items within each of the three divisions is broadly similar.

Art and Architecture

|  | Unique | % | Held by 2 | % | Held by 3 | % | Held by 4 | % | Total |
|---|---|---|---|---|---|---|---|---|---|
| CARM | 1,996 | 57.6 | 835 | 24.1 | 472 | 13.6 | 161 | 4.6 | 3,464 |
| UM | 43,074 | 55.0 | 26,318 | 33.6 | 8,775 | 11.2 | 161 | 0.2 | 78,328 |
| Mon | 23,960 | 44.9 | 20,456 | 38.3 | 8,775 | 16.4 | 161 | 0.3 | 53,352 |
| NLA | 50,708 | 66.1 | 16,985 | 22.1 | 8,882 | 11.5 | 161 | 0.2 | 76,736 |
| *Totals* | | | | | | | | | |
| *Holdings* | 119,738 | 56.5 | 64,594 | 30.5 | 26,904 | 12.7 | 644 | 0.3 | 211,880 |
| *Items* | 119,738 | 74.3 | 32,297 | 20.0 | 8968 | 5.6 | 161 | 0.1 | 161,164 |

Table 2:  Unique holdings and overlap for Art and Architecture division

The 161,164 Art and Architecture items have a total of 211,880 holdings, with an average of 1.31 holdings per item.

It is notable that the percentages of unique items held are very similar within other major humanities subject divisions. For example, for the division 'Language, Linguistics and Literacy', results for uniqueness for the three library collections were University of Melbourne, 54.4%; Monash University, 44.5%; and the National Library, 63.5%.

Sociology

|  | Unique | % | Held by 2 | % | Held by 3 | % | Held by 4 | % | Total |
|---|---|---|---|---|---|---|---|---|---|
| CARM | 2,659 | 40.0 | 1,682 | 25.2 | 1,540 | 23.1 | 773 | 11.6 | 6,654 |
| UM | 21,907 | 33.5 | 24,319 | 37.2 | 18,422 | 28.2 | 773 | 1.2 | 65,421 |
| Mon | 17,704 | 29.1 | 23,500 | 38.6 | 18,930 | 31.1 | 773 | 1.3 | 60,907 |
| NLA | 63,450 | 60.2 | 22,275 | 21.1 | 18,981 | 17.9 | 773 | 0.7 | 105,479 |
| *Totals* | | | | | | | | | |
| *Holdings* | 105,720 | 44.3 | 71,776 | 30.1 | 57,873 | 24.3 | 3092 | 1.3 | 238,461 |
| *Items* | 105,720 | 65.4 | 35,888 | 22.2 | 19,291 | 11.9 | 773 | 0.5 | 161,672 |

Table 3:  Unique holdings and overlap for Sociology division

The 161,672 Sociology items have a total of 238,461 holdings, with an average of 1.47 holdings per item.

Medicine

|  | Unique | % | Held by 2 | % | Held by 3 | % | Held by 4 | % | Total |
|---|---|---|---|---|---|---|---|---|---|
| CARM | 3,296 | 46.9 | 2,002 | 28.5 | 1,298 | 18.5 | 434 | 6.2 | 7,030 |
| UM | 34,977 | 44.9 | 29,206 | 37.5 | 13,269 | 17.0 | 434 | 0.6 | 77,886 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Mon | 39,843 | 49.0 | 27,556 | 33.9 | 13,562 | 16.7 | 434 | 0.5 | 81,395 |
| NLA | 47,601 | 56.9 | 22,136 | 26.4 | 13,559 | 16.2 | 434 | 0.5 | 83,730 |
| *Totals* | | | | | | | | | |
| *Holdings* | 125,717 | 50.3 | 80,900 | 32.3 | 41,688 | 16.7 | 1,736 | 0.7 | 250,041 |
| *Items* | 125,717 | 69.7 | 40,450 | 22.4 | 13,896 | 7.7 | 434 | 2.4 | 180,497 |

Table 4:  Unique holdings and overlap for Medicine division

The 180,497 Medicine items have a total of 250,041 holdings, with an average of 1.36 holdings per item.

Observations relating to the data

The NLA has the most recorded holdings for 18 of the 24 divisions. The exceptions are Art and Architecture (see Table 1), Chemistry, Computer Science, Mathematics, Music, and Physical Science. It is also the case that for 23 of the 24 divisions the NLA recorded the highest percentage of unique items, usually by a considerable margin (the exception was Library Science). As discussed above, this can be explained by the nature (breadth) of their collecting. It is also likely, however, that the degree of uniqueness in a collection has some correlation with collection size. This is apparent when comparing results for the two academic libraries. For 23 of the 24 divisions the larger of the two collections was also the one that recorded the higher percentage of unique items. This can logically be explained in that smaller collections will be driven by the need to acquire a core set of curriculum driven items, with a greater likelihood of duplication in other collections. As collections become larger they will inevitably focus on more research-related material, with a corresponding decline in duplication. The one exception was again Library Science, where The University of Melbourne has a slightly smaller collection than Monash University, but a higher percentage of unique items. This is almost certainly explained by the fact that the University of Melbourne collection has been developed for use by library staff rather than to serve a curriculum (the university does not educate in the area of library and information studies).

Tables 2-4 reveal a considerable difference in the results for the sample disciplines represented. The difference in results between Art and Architecture (humanities) and Sociology (social sciences) indicate the substantially higher level of uniqueness and lower rate of duplication (as indicated by average holdings per item) of the former. The results do not, however, suggest there is a linear progression from humanities to sciences, as Medicine has produced an outcome that is placed between these two extremes. There is evidence from other Conspectus divisions indicating that the humanities tend to produce a lower level of overlap than other discipline areas, but this requires closer examination.

**Discussion**

One of the challenges inherent in overlap studies is the interpretation of the results. There are no benchmarks available for assessing a 'high', 'low', or 'acceptable' level of overlap. Establishing an acceptable level of overlap is particularly difficult when, as in this case, there are no cooperative collecting agreements in place designed to minimise duplication and overlap. When libraries are driven by the needs of curricula—as in the case of the two university libraries—or by commitments to comprehensive collecting—as in the case of the National Library—then a degree of overlap is both unavoidable and necessary. It is also the case, however, that in a nationwide network of research libraries where efficiency in collection storage is at a premium, that reduced long-term overlap in the retention and storage of low-use print material will benefit the system as a whole. These benefits in turn have the potential to flow through to further efficiencies in the discovery and delivery of research materials in a system where a repository such as the CARM Centre has a commitment to permanent retention of low-use material in a high density storage environment. The National Library is also obligated to the permanent retention of Australian material.

The presence of in excess of 1.1 million duplicate holdings for the collections studied is indicative of the potential for de-duplication. Obviously this overlap number would grow—and grow quite quickly—as additional libraries were added to the calculation. The National Library has 666,026 duplicated holdings within this small sample of the academic library sector alone. It is of course the case that many of these will be part of the National Library's Australiana collections, but there is nonetheless scope for a more intensive examination of the characteristics of this duplicated material.

Further insight into the extent of the overlap can be gained by examining additional data recording the duplication between collections. Table 5 reports the overlap for *all* of the recorded holdings on WorldCat—as opposed to the 24 divisions in Table 1—for the two university-based collections included in this study, and the National Library.

|  | UM | Mon | NLA |
|---|---|---|---|
| UM | 1,524,110 | 482,845 | 520,052 |
| Mon | 482,845 | 1,405,960 | 547,486 |
| NLA | 520,052 | 547,486 | 3,233,921 |

Table 5: Three-way overlap, University of Melbourne, Monash University, and National Library.

Of note in these results is that 34.3% of the Monash University collection is duplicated by the University of Melbourne, and that 31.6% of the University of Melbourne collection is duplicated by Monash. In addition both academic libraries have considerable duplication with the National Library; 34.2% in the case of University of Melbourne, and 38.9% for Monash. The data in Table 5 again indicate

that there is considerable potential for de-duplication, but the exact extent of possible de-duplication can only be confirmed by closer examination of the items that comprise the overlap. This would be necessary in order to establish the characteristics of duplicated items and whether there is likely to be ongoing demand for this material that would justify retention in more than one library, on in main library sites as opposed to storage.

With access to a service such as *WorldCat Collection Analysis* it should be feasible to undertake this additional level of analysis. Although not utilised in the present study, *WorldCat Collection Analysis* offers access to more detailed levels of data regarding collection overlap. This includes the capacity to collate and compare holdings by features such as publication date, format and audience level. Also of particular relevance to the issue of understanding overlap and unique holdings in the national context is the capacity to establish 'groups' of libraries for comparison purposes. This might include, for example, groups that represent sectors within the university library community, such as the Group of Eight (research intensive) or the ATN (technology based) university libraries; or groups from outside the university sector such as major special libraries, or the Technical and Further Education libraries . Collection comparisons can then be made within, or between these groups, with a view to assisting collection management decisions of either individual libraries; the particular group or network to which they belong; or to the wider library community. Indeed it is only when non-university based libraries or groups of libraries are included in the data-gathering that a full picture of the collective wealth of Australia's research libraries will emerge.

While data related to overlap and uniqueness has previously been available in Australia from the NBD, it has been extremely difficult to mine, with no National Library service or software function specifically designed to meet the need. It has therefore not been possible for Australian libraries to use the NBD data in order to optimise its potential to assist in managing either local or system-wide collections. *WorldCat Collection Analysis* software, however, uses *the* major international union catalogue to make possible a rapid and detailed analysis of local collections, and to enable libraries to undertake collection comparisons on a scale of their own choosing. This extends to providing Australian libraries with the capacity to benchmark using international collections.

The breadth of the coverage of the WorldCat database, supported by of *WorldCat Collection Analysis* software, also provides an opportunity to broaden the basis for conceptualising and managing Australia's national research collection. The comparative ease with which collection analysis can be undertaken using *WorldCat Collection Analysis* makes it conceivable to include a wider range of collections within the scope of overlap-based studies, and therefore within any framework for collaborative planning and management of the national research collection. For while there has been acknowledgement that special libraries include valuable research material that is unlikely to be duplicated in academic libraries (Stephens 2009), the

practical difficulty of including these libraries within any data collection exercises has meant that they have been largely excluded. This exclusion, for example, has extended to the recent Australian overlap studies mentioned above.

**Conclusion**

The results of this pilot study add to the growing body of data regarding the potential for the rationalisation of print storage in ways that might produce benefits for the Australian research collections. It is apparent, however, that the data available as yet is preliminary and partial and that a much more complete investigation of both unique holdings and overlap are required.

The study has also identified that there are ongoing problems in the accuracy of some Australian holdings data in WorldCat, but that nonetheless the WorldCat database and its collection evaluation software have the potential to provide important data in support of the management of Australian research collections. It is also possible to conclude that the *WorldCat Collection Analysis* software is appropriate for the subsequent and expanded phases of this research, and that it is also likely to have substantial benefits for other Australian libraries interested in a better understanding of their collections.

**Acknowledgement**

**References**
Australian Research Libraries Collection Analysis Project, *Report* 2004. Available at http://www.library.uwa.edu.au/__data/assets/pdf_file/0004/524794/arlcap_final_report.pdf

Beals, Jennifer, and Ron Gilmour. 2006. Assessing collections using brief tests and WorldCat Collection Analysis. *Collection Building* 26(4): 104–107.

Connaway, Lynn S., Edward T. O'Neill, and C. Prabha. 2006. Last copies: what's at risk? *College & Research Libraries* 67(4): 370-379.

Genoni, Paul, and Eva Varga. 2009. Assessing the potential for a national print repository: results of an Australian overlap study. *College & Research Libraries* 70(6): 555-567.

Gherman, Paul. 2007. The North Atlantic Storage Trust: maximising space, preserving collections. *Portal: Libraries and the Academy 7*(3):273-275.

Lavoie, Brian, Lorcan Dempsey, and Lynn S. Connaway. 2006. Making data work harder. *Library Journal* 131(1). Available at
 http://www.libraryjournal.com/article/CA6298444.html

McClure, Jennifer Z. 2009. Collection assessment through WorldCat. *Collection Management* 34(2): 79-93.

Missingham, Roxanne and Robert Walls. 2003. Australian university libraries: collections overlap study. *Australian Library Journal* 52(3): 247-260.

National Library of Australia. OCLC Agreement. Available at http://www.nla.gov.au/librariesaustralia/oclc.html

OCLC: Introduction to the WorldCat Collection Analysis service, http://www.oclc.se/support/documentation/collectionanalysis/using/introduction/introduction.htm

O'Connor, Steve, and Cathie Jilovsky. 2008. Approaches to the storage of low use and last copy research materials. *Library Collections, Acquisitions & Technical Services* 32 (3/4): 121-126.

O'Neill, Edward T., Lynn S. Connaway, and Timothy J. Dickey. 2008. Estimating the audience level for library resources. *Journal of the American Society for Information Science and Technology* 59(13): 2042-2050.

Orcutt, Darby, and Tracy Powell. 2006. Reflections on the OCLC WorldCat Collection Analysis tool: we still need the next step. *Against the Grain* 18(4):44–48.

Payne, Lizanne. 2007. *Library Storage Facilities and the Future of Print Collections in North America* (Dublin, Oh: OCLC, 2007): 5. Available at http://www.oclc.org/programs/publications/reports/2007-01.pdf.

Rochester, Maxine K. 1987. The ABN database: sampling strategies for collection overlap studies. *Information Technology and Libraries* 6(3): 190-199.

Sharp, Steven. 2009. No more room aboard the ark! A UK higher education perspective on space management. *Interlending & Document Supply* 37(3): 126-131.

Stephens, Matthew. 2009. Heritage book collections in Australian libraries: what are they, where are they and why should we care? *Australian Library Journal* 58(2):173-189.

Vattulainen, Pentii. 2004. National repository initiatives in Europe. *Library Collections, Acquisitions, & Technical* Services 28 (2): 39-50.

Ward, Suzanne M., and Mary C. Aagard. 2008. The dark side of collection management: deselecting serials from a research library's storage facility using WorldCat collection analysis. *Collection Management* 33(4): 272-287.

White, Howard D. 2008. Better than brief tests: coverage power tests of collection strength. *College & Research Libraries* 69(2): 155-174.

Yoon, Hee-Yoon, and Sun-Kyung Oh. 2008. Shortage of storage space in Korean libraries: solutions centering upon hub-based collaborative repositories. *Aslib Proceedings* 60(3): 265-282.

Associate Professor Paul Genoni teaches with the Department of Information Studies at Curtin University in Perth. He has published numerous papers related to collection management, reference services and continuing professional development. From 2004 to 2010 he served as an educator representative on the ALIA Education Reference Group, and the Education and Workforce Planning Standing Committee.

Janette Wright, Chief Executive of CAVAL Ltd, has extensive management experience within the library and information services sector. Previous roles include Director, RMIT Publishing, aggregator and online publisher of Australian scholarly content, and Managing Director, of the journal subscription agency, RoweCom Australia/Divine Information Services - Europe. As Director of Public Library and Network Services at the State Library of NSW and Director, Library and Community Services, at Waverley Council, NSW, Janette has managed significant operational and grant funding programs for libraries with accountability at a senior level across a number of different legislative environments.