Advisory Committee Meeting

Friday, 8 April 2011
10.00 am – 3.45 pm
Council Room
National Library of Australia
LIBRARIES AUSTRALIA ADVISORY COMMITTEE
MEETING

Council Room, National Library of Australia
Friday 8 April 2011

DRAFT AGENDA

10:00 am Coffee on arrival
1. Welcome and introduction, apologies
2. Draft Minutes of the previous meeting and business arising LAAC/2011/1/01
3. Libraries Australia Status Report LAAC/2011/1/02
4. Libraries Australia statistics LAAC/2011/1/03
5. Report from the National Library (oral)

11.15 am Morning tea

FOR DISCUSSION
6. Report from the OCLC Asia Pacific Regional Council (oral)
7. Subscription model changes for 2011 LAAC/2011/1/04
8. Changes to the Libraries Australia User Agreement 2011 LAAC/2011/1/05

12.45 pm Lunch
9. Trove Status Report LAAC/2011/1/06
10. Revision of the Resource Sharing & Innovation Division Strategic Plan LAAC/2011/1/07

FOR INFORMATION
11. Evaluation of the 2010 Annual Forum LAAC/2011/1/08
12. Reimagining Libraries’ Project Reports (oral)
14. Other business
15. Conclusion & Review of resolutions (if any)

3.45 pm Close of Meeting
DRAFT MINUTES OF THE LIBRARIES AUSTRALIA ADVISORY COMMITTEE

National Library, Canberra
Tuesday 19 October 2010

<table>
<thead>
<tr>
<th>Committee Members Present:</th>
<th>Officers Present:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chair: Ms Anne Horn</td>
<td>Dr Warwick Cathro</td>
</tr>
<tr>
<td>University Librarian</td>
<td>Acting Director General</td>
</tr>
<tr>
<td>Deakin University</td>
<td>National Library of Australia</td>
</tr>
<tr>
<td>Ms Liz Burke</td>
<td>Ms Debbie Campbell</td>
</tr>
<tr>
<td>Director</td>
<td>Acting Assistant Director-General</td>
</tr>
<tr>
<td>Library Services</td>
<td>Resource Sharing and Innovation</td>
</tr>
<tr>
<td>Murdoch University</td>
<td>National Library of Australia</td>
</tr>
<tr>
<td>Mr Vic Elliott</td>
<td>Ms Laurel Paton</td>
</tr>
<tr>
<td>Director, Scholarly Information Services &amp; University Librarian</td>
<td>Acting Director, Collaborative Services</td>
</tr>
<tr>
<td>Australian National University</td>
<td>National Library of Australia</td>
</tr>
<tr>
<td>Ms Pam Gatenby</td>
<td>Mr Rob Walls</td>
</tr>
<tr>
<td>Assistant Director-General</td>
<td>Director, Database Services</td>
</tr>
<tr>
<td>Collections Management</td>
<td>National Library of Australia</td>
</tr>
<tr>
<td>National Library of Australia</td>
<td></td>
</tr>
<tr>
<td>Ms Karen Hansen</td>
<td>Ms Helen Ludellen</td>
</tr>
<tr>
<td>Section Manager, Library Services</td>
<td>Manager, Australian National Bibliographic Database</td>
</tr>
<tr>
<td>Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA)</td>
<td>National Library of Australia</td>
</tr>
<tr>
<td>Ms Noelle Nelson</td>
<td>Minutes: Ms Mary-Louise Weight</td>
</tr>
<tr>
<td>Director, Library Services</td>
<td>Libraries Australia Customer Services</td>
</tr>
<tr>
<td>State Library of New South Wales</td>
<td>National Library of Australia</td>
</tr>
<tr>
<td>Ms Ann Ritchie</td>
<td></td>
</tr>
<tr>
<td>Editor</td>
<td></td>
</tr>
<tr>
<td>Australian Library Journal</td>
<td></td>
</tr>
<tr>
<td>Mr Geoff Strempel</td>
<td></td>
</tr>
<tr>
<td>Associate Director, Public Library Services, South Australia</td>
<td></td>
</tr>
<tr>
<td>Ms Monika Szunejko</td>
<td></td>
</tr>
<tr>
<td>Manager, Access</td>
<td></td>
</tr>
<tr>
<td>State Library of Western Australia</td>
<td></td>
</tr>
</tbody>
</table>

A summary of action items is included in Attachment A.

The meeting commenced at 10:30 a.m.
Agenda Item 1
Welcome, Introductions and Apologies

Ms Anne Horn welcomed all present to the meeting. Apologies were received from Mr Chris Taylor, Executive Manager, Library Resources and Technology Services, University of Queensland.

Ms Horn noted that Dr Warwick Cathro was currently Acting Director General; Ms Debbie Campbell was Acting Assistant Director General, Resource Sharing and Innovation and Ms Laurel Paton was Acting Director, Collaborative Services. Dr Warwick Cathro joined the meeting after attending a Senate Estimates Committee hearing.

Ms Horn advised that Ms Karen Hansen, Section Manager, Library Services, Department of Families, Housing, Community Services and Indigenous Affairs would be leaving the Committee and thanked her for her contribution.

Ms Horn paid tribute to retiring Director-General Ms Jan Fullerton and acknowledged her significant contribution to the National Library and the profession.

Agenda Item 2
Draft Minutes of the Previous Meeting and Business Arising
LAAC/2010/2/1

The Committee resolved to accept the minutes of the previous Committee meeting.

Ms Campbell and Mr Rob Walls reported on the actions arising from the previous Committee meeting.

Agenda Item 2
1. Ms Campbell confirmed that the subscription model for university sector was reviewed with assistance from Ms Horn as LAAC chairperson. The subscription model for state and territory libraries and public libraries will be reviewed in early 2011.

2. Ms Campbell noted that the best approach with regards to data contribution is to liaise with different contributors to give specific advice. If contributors are already members of Libraries Australia, this would be the preferred method of data contribution. Other contribution pathways were noted and Ms Campbell advised that advice the contribution pathways may change over time as more become available.

3. Mr Walls reported that 806 Libraries Australia members have some holdings on the Australian National Bibliographic Database (ANBD) and 400 have none. 540 libraries have updated or added to their holdings on the ANBD in the last 12 months. Libraries not maintaining holdings will be contacted. Mr Elliott confirmed that this is an issue for WorldCat membership as well. Ms Campbell added that additional information will be provided in the Libraries Australia User Agreement about this and the benefits of WorldCat membership.

Agenda Item 5
The Resource Sharing & Innovation Division’s Strategic Plan was made available to Committee members for comment in July.
A paper on the progress of the ARDC Party Infrastructure Project is to be tabled at today’s meeting.

Ms Rose Holley will speak about web 2.0 research at the annual Libraries Australia Forum 2010.

**Agenda Item 7**
Libraries Australia has considered the matter of TAFE consortia subscriptions further and has decided to continue with the original proposal.

**Agenda Item 10**
With regard to establishing interoperability between WorldCat Resource Sharing and Libraries Australia Document Delivery, there has been a further development which will be discussed under Other Business.

**Agenda Item 11**
Ms Campbell advised that the Trove marketing plan is not ready yet – another area of the NLA will complete it in conjunction with the NSLA Marketing Working Group. The plan will be finalised by December and distributed then.

Following a discussion with Dr Cathro at the last Committee meeting regarding the low visibility of special library collections, Ms Ritchie and Ms Hansen commenced a project to address the issue. They analysed the results of the ANBD coverage survey done in 2008/9 which asked why libraries were not participating fully in putting collections on the ANBD. The reasons given were the nature of the collections concerned, issues of local resourcing and an assumption that the materials would have to be available for loan. Ms Ritchie will give a presentation on the findings at the Libraries Australia Forum.

Project work will continue, and two national studies which are underway will assist with this. Ms Campbell advised she will be speaking at the next ALIES Conference (Australasian Libraries in the Emergency Sector) in April 2011. Ms Hansen added that a paper including suggestions and recommendations will be produced at the end of the project.

The request for comparative search statistics of Libraries Australia and Trove has been satisfied by their inclusion in the Libraries Australia Annual Report 2009/10.

**Agenda Item 3**
Libraries Australia Annual Report 2009/2010
LAAC/2010/2/2

Ms Campbell noted the steady growth in the service. Some blog posts from last year’s Forum and the different ways Libraries Australia uses to keep in touch with members are detailed in the report. A small number of cancellations has been offset by some growth in membership. The statistics for use of the various databases provided through Libraries Australia show that the ANBD remains the most heavily used.

Mr Walls advised that NBD growth exceeded one million bibliographic records and one million holdings. The net growth in holdings is lower than last year which reflects the holdings updates which have occurred. The addition of OCLC identifiers to the ANBD has been completed, realising two benefits: more deep linking to local library systems and improved record matching.
Mr Walls expanded on the duplicate detection work which has been undertaken so far by configuring the software to run using ISBN numbers as match keys, removing 634,000 records from the ANBD. It will next be done using OCLC identifiers and then using author/title keys. Ms Horn commented that the duplicate detection work should be celebrated.

Mr Walls talked about the current work to develop a batch service for uploading authority data. Libraries Australia is working with OCLC to automate this process.

The work done this year in global holdings update was noted by Mr Walls, whereby around two million holdings have been updated for 80 libraries. There was some discussion about statistical fluctuations in search activity and various reasons were given for this such as changes in workflows, e-records and specific projects.

Ms Horn noted the opportunity for improving the linkages between schools and universities and the opportunity for university libraries to publicise Trove more fully, for which marketing assistance would be very helpful. Ms Campbell advised that a letter had been sent to all schools about Trove, but more materials could be made available.

The Committee noted the report.

**Agenda Item 4**

**Libraries Australia Statistics**

LAAC/2010/2/3

Ms Campbell explained that the difference between expenses and revenue represented the National Library’s subsidy of Libraries Australia. There was some discussion about this and Ms Campbell assured the Committee that the level of subsidy will stay as it is for now. The slightly lower revenue can be explained by the switch in subscription models. She also noted the effect of the University of Queensland holdings refresh on the statistics.

The Committee noted the report.

**Agenda Item 5**

**Libraries Australia Activity Report July – October 2010**

LAAC/2010/2/4

Mr Walls added to what had been discussed in the Annual Report by mentioning the recent completion of the Library of Congress Table of Contents (TOC) data project. TOC data has been added to 240,000 records. TOC data is now harvested regularly.

There has been follow up on the 2008/09 ANBD coverage survey. Several state and university libraries have been visited by Libraries Australia staff and this has resulted in additional contributions to the ANBD, of note a contribution of 40,000 records from Murdoch University. It is important to communicate the minimum record standard, as a modest level of cataloguing should not be seen as a disincentive to contribution. The next focus for addition will be on records of large microform collections.

The recent CBS upgrade included additional support for SRU Record Update. An OCLC project is underway to transmit higher level record enhancement data from WorldCat to Libraries Australia as a way of adding further value. SRU Record Update will be used as an
alternative to the Record Import Service for loading data from Australian libraries in future.

Mr Walls commented on the work being done by OCLC with regard to the WorldCat Synchronization Gateway and the implications of this for Australian libraries.

Currently 12 libraries have agreed to participate in a project which is an attempt to develop a service to keep holdings up to date. Libraries with less than 200,000 holdings were contacted and asked to send their holdings via RIS as a possible way of working out holdings which are no longer current.

Ms Horn commented on the value of deep linking and suggested it could be useful in testimonials to gather feedback on this. Ms Paton advised that 211 libraries are now deep linked and acknowledged the positive feedback already received.

Mr Walls advised that work is being done to develop a program which automatically links bibliographic records with related authority records. Work is also being done to enhance the quality of authority records in Libraries Australia by using data coming from the Virtual International Authority File (VIAF). Work is also being done with OCLC to add detailed holdings statements for serials to WorldCat.

Mr Walls advised that Libraries Australia is moving closer to an agreement with the British Library Document Supply Centre with regard to interoperating with Libraries Australia Document Delivery (LADD).

Ms Paton drew attention to the statistical snapshots which are now available to any library which requests them. Libraries can also access a snapshot of their search activity at www.nla.gov.au/librariesaustralia/statistics/index.html.

Ms Paton noted that over 230 delegates were registered for the Libraries Australia Forum. New promotional items were to be distributed – a USB flash drive, and an eco-bag.

To assess the longterm effect of the roadshow held in 2009, 25 libraries were contacted and seven agreed to be interviewed. They gave very positive feedback, although 50% said that their workflows hadn’t changed. The difficulty of engaging CEOs and senior managers in roadshows and the Forum was discussed.

Ms Burke suggested attaching a meeting to a scheduled CAUL meeting as a way of engaging with senior managers. Ms Campbell advised that she would be giving a presentation at the NSLA CEOs meeting in November. It was agreed that a meeting attached to CAUL would need to be strategic, well structured and high level and could include other sectors apart from CAUL members. The meeting should dispel myths, look to the future and examine the relationship between Libraries Australia and OCLC. Ms Campbell noted that the LAAC papers are made public, but this was thought to be too retrospective. The Committee expressed its interest in assisting with communicating with the Libraries Australia community.

The Committee noted the report.

**Agenda Item 6**

**Report from the National Library (oral)**

Dr Cathro advised that the National Library’s new minister is Mr Simon Crean, and the National Library is now part of the Prime Minister and Cabinet portfolio and the Deputy Secretary is Mr Richard Eccles. The new chair of the National Library Council is the
Honourable New South Wales Chief Justice James Spigelman. He has taken a strong interest in the National Library’s digitisation program and digital legal deposit issues.

There have been several farewells for Ms Jan Fullerton who officially retires on 9 November. The selection process for the next Director General is underway.

Dr Cathro attended the OCLC Asia Pacific Council meeting in Tokyo on 6-7 September. He gave a presentation about Trove and public engagement and visited the National Diet Library, noting in particular their digitisation program.

Dr Cathro gave an update on a meeting of the House of Representatives Standing Committee on Education & Training Inquiry into School Libraries and Teacher Librarians in Australian Schools at which he and Ms Roxanne Missingham discussed Electronic Resources Australia (ERA). This Committee lapsed when the election was called in July and it is not certain whether it will be re-instated.

Tenders have been called for the Treasures Gallery project now that the design phase is complete. It should open in late 2011.

Dr Cathro advised that the NLA has been assessing what to do with regard to the failure of the Digital Deluge policy proposal. A plan to meet essential minimum requirements for digitization and for replacing software needs to be developed. Different prototypes and approaches in this area are being examined.

Ms Gatenby advised that the Australian Women’s Weekly digitisation will be completed by the end of 2010, and content will be made available through the newspapers zone of Trove. A major national launch is being planned in conjunction with Australian Consolidated Press and Ms Deborah Thomas. The first phase of the Newspaper Digitisation Project will be completed by June 2011 and will include 93 titles covering all the major capital city dailies. Contributor guidelines are being distributed to public libraries. The preferred model is for the National Library to digitize for others with the costs paid for by the contributing library.

The National Library is collaborating with other National Libraries to develop a methodology for costing cataloguing as a basis for comparison. Costs seem to be very similar so far. The time cataloguers are spending on other activities, not just cataloguing, is a commonality. The results of this project will be made available when complete.

Ms Gatenby gave an update on Resource Description & Access (RDA) which was released in June 2010. The online toolkit is now available for subscription. The National Library will jointly implement when the Library of Congress, British Library and Library and Archives Canada do, which is unlikely before early 2012. The National Library is working on developing training materials and a network of trainers who will be trained in 2011.

The Committee noted the report.

Agenda Item 7
Resource Sharing and Innovation Division Strategic Plan
LAAC/2010/2/5

Ms Campbell thanked the Committee for comments which were integrated into the draft Strategic Plan July 2010 – June 2012. The next six monthly review will occur in January 2011. Ms Campbell advised the Committee of modifications to update the environmental scan and
further highlight the value proposition offered by WorldCat.

Ms Horn asked how Committee members can assist to extend and maintain partnerships with member libraries. Ms Campbell discussed some of the outreach activities which currently take place and added that she would welcome hearing more from Committee members about topical issues. Dr Cathro noted two issues Libraries Australia would like to distribute information about: making more timely contributions to the ANBD using SRU Record Update, and exposing hidden collections with collection level entries in the ANBD for indexing in Trove.

Ms Gatenby commented that while the indexing of full-text lists is very useful and beneficial for Trove, there are some major disadvantages for libraries contributing brief collection records, so it should be considered with that in mind. Libraries must read the guidelines first to confirm their capability and seek assistance if required.

Ms Gatenby asked if Libraries Australia still saw its role as purchasing further datasets and Mr Walls advised that there is no available money to purchase content above what is already being purchased. Mr Walls also commented that it is difficult to establish how useful particular datasets are. Ms Horn raised the issue of creating records for e-books and there was discussion about how Libraries Australia may be able to assist with this.

Ms Nelson requested further information about how the subscription model for state territory and public libraries may change. Ms Campbell advised that it is based on population data, so the original formula will be updated with the most recent ABS figures.

In response to Ms Szunejko’s question about the development of local views of Trove, Ms Campbell advised that in conjunction with Reimagining Libraries’ Project 2, the Trove team has been canvassing interest to establish what would be required in an Application Programming Interface (API) and is currently analysing the results of a survey. The API development work should be finished within 12 months.

Dr Cathro advised that Libraries Australia has renewed the agreement with OCLC for another four years and commented that only very slight changes were made.

Ms Ritchie sought clarification about whether special libraries would be included in the follow up to the ANBD coverage survey. Mr Walls confirmed that managers of special collections which have been identified across the sectors will be contacted; some of which are from special libraries.

Mr Strempel commented that Electronic Resources Australia (ERA) is not only of benefit to small libraries. Ms Horn confirmed that the ERA submission to the House of Representatives Inquiry into School Libraries and Teacher Librarians in Australian Schools is very relevant and it is to be hoped that it is reconstituted because of the focus on pathways to university. Dr Cathro confirmed that there will be a continuation of lobbying relevant government agencies with regard to funding national licences to e-resources.

Ms Campbell reported that Trove development is on schedule. There was discussion about a shift in the type of feedback being received. Users are becoming more accustomed to the service and awareness of the service is growing. Trove is now being used for research purposes.

The Committee noted the report.
ACTION: LAAC members to contact Ms Campbell with suggestions about the topical issues to discuss when making visits to university and state libraries; to generate more interest in Libraries Australia contact; to map state and sectoral connections, and to seek invitations for Libraries Australia staff to attend sectoral meetings.

ACTION: Ms Gatenby’s suggested wording changes with regard to the Reimagining Libraries Project 8 (Description and Cataloguing) to be incorporated into the Strategic Plan when it is reviewed.

Agenda Item 8
Report from the OCLC Asia Pacific Regional Council
LAAC/2010/2/6

Mr Elliott reported on the Asia Pacific Regional Council meeting held at the National Library in Canberra on 15 April 2010. There were 60 delegates. The meeting was well received and there was enthusiasm to have the meeting in Canberra again, possibly in 2012.

The Global Council meeting was held in Dublin, Ohio 19 – 22 April, 2010. There were presentations from the Regional Councils, a report from Mr Jay Jordan, OCLC CEO and from the Chair of the Board of Trustees. Other presentations covered the OCLC value proposition, an update on OCLC research, a look at the impact of mass digitization and an extended discussion on the Record Use Policy.

A virtual meeting of the Global Council was held on 21 June 2010 to vote on various documents including the membership and governance protocols; the statement on shared values and principles of the OCLC co-operative; and the recommendations of the cost sharing and pricing strategy taskforce, including setting up a new cost sharing taskforce to advise OCLC on a new model. The revised OCLC/WorldCat principles of cooperation were examined and approved.

A briefing was given on the new Record Use Policy which had previously been approved by the OCLC Board of Trustees on 14 June.

Mr Elliott drew the Committee’s attention to the report on the OCLC Asia Pacific Regional Council Membership Conference held in Tokyo on 6-7 September 2010. The 2011 meeting will be held in Taipei, Taiwan.

Mr Elliott advised that the next Global Council virtual meeting to be held on 25/26 October would look at Global Council activities for the forthcoming year. There will also be an in-person Global Council Executive meeting in Dublin, Ohio from 24 – 26 October. Mr Elliott advised that his term will finish on 30 June 2011.

Dr Cathro commented on his attendance at the Tokyo meeting, noting that all three rungs of the governance structure of OCLC were in attendance. There was further discussion about issues which apply to the region particularly and Mr Elliott confirmed that raising local issues is appropriate.

The Committee noted the report.

Ms Campbell introduced Ms Helen Ludellen who has replaced Ms Rajapatirana as Manager,
Australian National Bibliographic Database. Ms Ludellen attended the afternoon session of the Committee meeting.

**Agenda Item 9**  
**Draft ANBD Scope Statement**  
**LAAC/2010/2/7**

Mr Walls presented the scope statement as a response to discussion at the previous two LAAC meetings about the relationship between Libraries Australia and Trove, particularly regarding data contribution pathways. He noted that the proposed statement is not redefining the scope of the ANBD, rather it is a restatement and clarification of what the policy has always been. The scope statement aims to be more specific about different types of records.

Ms Szunejko commented and the Committee agreed that the new ANBD Scope Statement and its purpose were very clearly stated. Mr Walls commented that the document was not intended to address how the ANBD relates to Trove. There was further discussion about the relationship. Mr Walls confirmed that article level metadata for journals is out of scope for the ANBD but in scope for Trove.

In response to a question from Ms Horn seeking clarification on the juxtaposition between the ANBD, Libraries Australia and Trove, Dr Cathro clarified that the purpose of Libraries Australia is to support library workflows, Trove is for discovery of items in Australian collections including archives and museums, university repositories and that all of the ANBD data is also in Trove.

Ms Gatenby commented that the guidelines do not provide guidance on which service to contribute to and when, rather they are guidelines on what Libraries Australia will accept. The shortcomings for analytical data were noted. Ms Campbell noted that the Libraries Australia and Trove teams are available to discuss the most suitable data contribution pathways. Mr Walls added that it will be useful to review this document in a few years’ time when Libraries Australia has been migrated to the Lucene platform underpinning Trove.

RESOLUTION: The Committee resolved to accept the Scope Statement.

**Agenda Item 10**  
**Trove Report**  
**LAAC/2010/2/8**

Ms Campbell gave an update on Trove. There are now more than 101 million items in the service. The platform will be upgraded soon to manage the increased size. The Australian Newspapers interface will be shut down in November, and planning has commenced for the migration of other services starting with Australia Research Online, followed by Music Australia and then Picture Australia.

Ms Campbell explained how the information in the Australian Libraries Gateway is used in Trove and noted the importance of having this information up to date. The team will market this to member libraries. Ms Campbell explained how ALG data is sent to the WorldCat.
Registry. Although a manual process at this stage the team is working towards this becoming an automated process.

Ms Campbell highlighted the 23,000 visits to the site in a single day (Tuesday 18 October), the highest daily usage so far. Some additional usability testing was done, following on from a user survey which showed the profile of the users and the different enquiry types. The importance of the social engagement features of Trove is demonstrated by the user interaction activity statistics available from the Trove home page.

Ms Campbell mentioned the demand for training and the development of a series of instructional screencasts which are now available on YouTube. The positive public response to Trove was noted.

The Committee noted the report.

Agenda Item 11
Re-imagining Libraries Project 8 (Description and Cataloguing)
LAAC/2010/2/9

Ms Gatenby gave a report on the outcomes of the Re-Imagining Libraries’ Cataloguing and Description Project. The main purpose of the Project is to create awareness of collection lists which already exist within NSLA libraries that have not been made available online, and to find out what the barriers are to converting these lists from paper-based to machine readable data.

There are four proposed data models for improving national level access to information about collections that is contained in lists. The use of Trove functionality which allows indexing of lists has been enthusiastically received. Documentation of the models will be made publically available.

There was further discussion about progress which has been made by particular libraries including SLSA and SLNSW. Mr Walls commented that the issue is not what format the data is in, but the extent of the data and noted the limitations of Libraries Australia being MARC-based. Some of these limitations may be addressed when the Libraries Australia Search service is redeveloped on the Lucene platform underpinning Trove.

Dr Cathro expressed his interest in raising awareness of this and making this information available to other library sectors. It was noted that the issue would be addressed in a Special Libraries presentation at the Forum. There was further discussion about collections to target and ways of encouraging libraries to contribute.

The Committee noted the report.

Agenda Item 12
Reimagining Libraries Project 2 (Open Borders)
LAAC/2010/2/10

Ms Campbell gave a Project update. Work on stage 4 of Trove, to provide authenticated access to e-resources via Trove, has commenced. The team met with RMIT Publishing to arrange collection of subscription information and metadata for journal articles which will
be used as test data initially. Dr Cathro spoke about the benefits of the Project for e-resource vendors. Cengage Gale data as well as APAIS and AMI data will also be collected, and some of the more difficult authentication pathways will be tested by April 2011.

With regard to the other two aspects of the Project, Ms Campbell commented on the survey to establish what NSLA libraries are doing to push content into popular social spaces. She also talked about recent work to examine user annotations contributed to Trove and the possibility of extracting and sharing these across library systems.

The Committee noted the report.

**Reimagining Libraries Project 7 (Collaborative Collecting) - Oral report**

Ms Nelson gave an update on Project 7 which has now made substantial progress in identifying the core set of e-resources for subscription by all NSLA libraries. Ms Nelson thanked Ms Nikki Darby and Mr Rob Walls for their assistance. A report will go to the NSLA CEOs recommending the core set, a product list of 20, for which vendor pricing for the first 16 titles has been obtained. There will be a phased introduction.

The NSLA e-resources consortium Business Plan 2010 – 2012 is aligned to the work of Project 7. This includes timeframes for the phasing in of products, working with the vendors and the NSLA Marketing Working Group.

Ms Nelson also outlined the work of the Project to look at the concept of the last copy process and develop a draft agreement for considering retrospective collections, handling serials, identifying who would have the last copy and building complete sets. Originally assessed on a title basis, a subject basis is now being considered. The Project group is also examining existing policies and procedures for disposal.

Further work for the group came out of the NSLA Digital Preservation Forum to look at libraries’ collecting priorities. The group will do a survey to establish what needs to happen with regard to digital preservation. Ms Campbell asked whether the project would explore the potential offered by the Hathi Trust (to reduce local digitisation workloads).

There was further discussion about how to make the national subscriptions available and whether there would need to be any authentication for nationally licensed products. There was also further discussion about how to relate this work on last copy with university libraries’ holdings. The Project will start with NSLA libraries who will then relate to university holdings in their state and work to tidy up their holdings as the Project continues.

The Committee noted the report.

**Agenda Item 13**  
**The Resource Sharing Manifesto**  
LAAC/2010/2/11

Mr Walls drew the Committee’s attention to the Resource Sharing Manifesto which was adopted by the Rethinking Resource Sharing Initiative in 2007. The Committee discussed possible wider dissemination of the Manifesto.

The Committee recommended the Manifesto be shared more broadly with relevant Library
associations.

ACTION: The Committee to refer the Resource Sharing Manifesto to other groups for consideration.

Agenda Item 14
LAAC/2010/2/12
An overview of the Australian Research Data Commons Party Infrastructure Project (ARDCPIP)

Ms Campbell gave a brief summary of the project which is now halfway through its 18 month timeframe. A party persistent identifier scheme, in use in the people zone of Trove, has been used in the project. When the work is completed the results will be seen in the people zone of Trove. A party administration tool will be rolled out to the university sector in Stage 2. Given its technical complexity, an interface designer has been brought into the project to make sure the tool is as usable as possible.

Early implementers include Griffith University and Queensland University of Technology. Funding constraints mean between only five and 10 sites will be supported initially.

The achievement of upgrading the party infrastructure to support the Encoded Archival Context for Corporate bodies, Persons and Families (EAC-CPF) standard was noted, better aligning this with research sector requirements.

A presentation will be given at the upcoming CAIRSS Information Day in November by the Acting Project Manager Natasha Simons. There is further information about the project in the project wiki, http://wiki.nla.gov.au/display/ardcipp/Party+Data+Survey.

The Committee noted the report.

Agenda Item 15
Libraries Australia Forum and ACOC Seminar 2010 (oral)

The Forum was discussed under Agenda Item 5. Mr Walls gave a report on the ACOC seminar, scheduled for the day after the Forum, which attracted 213 registrations. An interesting programme of speakers included David Lindahl, Co-Executive Director of the eXtensible Catalog project as the keynote speaker and three short 20 minute presentations on innovative uses of metadata, RDA and other cataloguing standards.

The Committee noted the report.

Agenda Item 16
Article for information: “Libraries Australia and Trove: A special network for special libraries”

Ms Campbell introduced the article, a briefer version of which was published in August 2010 issue of Incite. Ms Ritchie mentioned a recent useful discussion which took place on the

**Agenda Item 17**

**Other Business**

1. **Membership of the Libraries Australia Advisory Committee**

   Ms Campbell advised that as Ms Hansen is stepping down from the Committee, a new representative will be sought for the non-elected position in time for the next meeting.

2. **WorldCat Resource Sharing (WCRS) agreement**

   Mr Walls advised that discussions have taken place between OCLC and Libraries Australia Document Delivery staff with regard to a WCRS agreement. A number of business and technical issues would need to be addressed.

   Dr Cathro asked the Committee to consider how Libraries Australia members would feel if the only way to use WCRS was via an International ILL Request Gateway fee. There was further discussion about this as a charging model.

   Mr Walls mentioned the survey done about the possibility of working with WCRS earlier in the year with LADD libraries. Of 100 responses about 76 libraries said they would be interested and 15 libraries said they already engaged in some form of international interlibrary loans. WCRS will be implemented on an opt in/opt out basis.

   Dr Cathro advised that Libraries Australia would continue discussions with OCLC.

**Conclusion and Review of Resolutions**

**RESOLUTIONS**

1. The Committee resolved to accept the minutes of the meeting held on 14 April 2010.
2. The Committee resolved to accept the updated ANBD Scope Statement.

Ms Campbell made a presentation to Ms Hansen and thanked her for her contribution to the Committee.

The meeting closed at 3.45pm.
Attachment A: Actions

Agenda Item 7
Resource Sharing and Innovation Division Strategic Plan
LAAC/2010/2/5

ACTION: LAAC members to contact Ms Campbell with suggestions about the topical issues to discuss when making visits to university and state libraries; to generate more interest in Libraries Australia contact; to map state and sectoral connections, and to seek invitations for Libraries Australia staff to attend sectoral meetings.

ACTION: Ms Gatenby’s suggested wording changes with regard to the Reimagining Libraries Project 8 (Description and Cataloguing) to be incorporated into the Strategic Plan when it is reviewed.

Agenda Item 13
The Resource Sharing Manifesto
LAAC/2010/2/11

ACTION: The Committee to refer the Resource Sharing Manifesto to other relevant groups for consideration.
LIBRARIES AUSTRALIA ACHIEVEMENTS
NOVEMBER 2010 – MARCH 2011

Summary

This progress report summarises achievements since the last briefing provided at the Libraries Australia Annual Forum on 20 October 2010 in Canberra¹.

The key achievements have been:

- the addition of the 100th Libraries Australia Document Delivery site using the ISO ILL protocol;
- the commencement of a review of the subscription model for public, state and territory libraries; international agencies and the User Agreements;
- the trialling of a de-duplication tool that identifies authority records in incoming files which duplicate existing ANBD authority records;
- the trialling of a new service that will make it easier for smaller libraries to update their ANBD holdings. It specifically aims to address the problem caused by the fact that some library management systems are unable to report deleted or updated holdings;
- the migration of two key software modules – the Libraries Australia CBS cataloguing service and Libraries Australia Document Delivery - to new servers in order to expand service capacity; and
- the commissioning of a new design for the information website.

Libraries Australia Cataloguing Service

Libraries Australia Cataloguing (CBS)

The Libraries Australia Cataloguing production service was moved to a new server on 3 December and the Training service was moved on 20 January. This was part of the National Library’s ongoing hardware replacement program. Extensive testing was performed prior to the moves to ensure that there was no disruption to service.

Record Import Service

A global refresh involving 128,721 records for Northern Territory Library (XNLS) was successfully completed. This was the first global refresh using the new Record Import Service stream, RIS36. RIS36 was implemented in 2010 to reduce the number of records sent to human review as a result of large holdings refresh loads. However RIS36 is sending very large numbers of records to review simply because of the quantity involved in refreshing a library’s holdings. As a result the Libraries Australia Database Services team has suspended global refreshes while it identifies strategies for reducing the amount of reviewing created by RIS36 in particular.

A large holdings refresh was also completed for the Parliament of Australia: Parliamentary Library.

A trial load of new authority records from the University of Adelaide (SUA:AUTH) using a new de-duplication tool has commenced. The de-duplication tool has been developed in-house to identify corporate name authority records in incoming files that duplicate existing ANBD corporate name authority records so that that these duplicates won’t be added to the ANBD. One file of new name authorities from the University of Adelaide has been loaded and work has commenced on a second file.

A list of the libraries that began using the Record Import Service during the period 1 November 2010 to 28 February 2011 is provided at Attachment B.

Maintenance of the Australian National Bibliographic Database

Between 1 July 2010 and 28 February 2011, Global Holdings Updates were performed for 123 libraries resulting in the updating, transfer or deletion of over 2.6 million holdings.

Holdings Deletion Trial

The following libraries are now involved in the holdings deletion trial aimed at making it easier for smaller libraries to update their ANBD holdings:

- Wyndham City Council Library Service: Werribee Library [VWYN]
- Box Hill Institute of TAFE Library [VBHE]
- Singleton Shire Council: Singleton Public Library [NSIN]
- SA Water Library [SEW]
- Corrective Services NSW: Library & Information Service [NCOR] and
- Mackay Regional Council - Mackay Libraries [QMC].

Early indications are that the new process will allow Libraries Australia to deduce which ANBD holdings are to be deleted or updated based on a file of all holdings sent by the contributing library. However, the success of this method of maintaining holdings depends on the quality of the data being supplied by the contributor and early indications suggest that this will only be effective where high quality data is supplied.
Record Export Service

Between 1 July 2010 and 28 February 2011, Libraries Australia customers were supplied with seven Library of Congress Table of Contents product files containing a total of 21,787 enriched records. Three separate concordance files were also supplied, containing a total of 1,905,257 listings that provided current and superseded ANBD numbers with their corresponding local system numbers.

SRU Record Update

Libraries Australia has been monitoring OCLC’s development of an SRU synchronisation gateway. Testing of the gateway by OCLC is underway with a number of library systems including Aleph, Amlib, Millennium and Voyager. This testing has included the transmission of records to a copy of the Libraries Australia Cataloguing system that is installed on a server at OCLC in Leiden. The Libraries Australia Database Services team plans to establish pilot sites in Australia later in the year.

Libraries Australia Document Delivery Service

LADD Development

During the period the Libraries Australia Database Services team has continued to negotiate with OCLC on the interconnection of Libraries Australia Document Delivery (LADD) and WorldCat Resource Sharing (WCRS).

Libraries Australia and the British Library Document Supply Centre (BLDSC) have reached agreement that will allow LADD users to request resources from the BLDSC. Interoperability testing has been performed and only minor issues identified. The final version of the BLDSC Contract has not been received.

Server migration

The LADD production system was moved to a new server on 18 March. This was part of the National Library’s ongoing hardware replacement program.

ISO ILL Protocol

A major milestone was reached on 11 November 2010 when Hobsons Bay Libraries became the 100th library to interconnect its ILL system with LADD using the ISO ILL protocol. Hobsons Bay Libraries, which uses the VDX system, is part of the Library Link Victoria consortium.

The table below shows the growth in the number of libraries interoperating with LADD using the ISO ILL protocol since 2003:
A list of the libraries which joined the Libraries Australia Document Delivery service during the period from 1 November 2010 to 28 February 2011 is provided at Attachment C.

On 28 February, a three month project commenced to review the options for the continuing operation and development of LADD including the strategic direction of the VDX software and the future of the ISO ILL protocol.

Libraries Australia Search

The searching activity in Libraries Australia has remained strong:

Graph 1: Growth in the number of sites which have implemented the ISO ILL protocol

Graph 2: Search activity on Libraries Australia
Relations with the Libraries Australia Membership

Memberships

There are 1,283 current members of Libraries Australia, an increase of 20 memberships since July. This is reflected in the increased number of Libraries Australia userids:

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>contributing libraries</td>
<td>6,023</td>
</tr>
<tr>
<td>non-contributing libraries</td>
<td>1,469</td>
</tr>
<tr>
<td>Total this year (last year 7,119)</td>
<td>7,492</td>
</tr>
</tbody>
</table>

Between November 2010 and February 2011 there were 10 cancellations, including one due to library closures and three individual accounts. In the same timeframe, there were 23 new registrations, including for four schools and four individuals. There are now 44 school library members.

By the end of February, 330 people had registered to use the Libraries Australia discussion space in ning.

Enquiries

In late October 2009, the software used by the Libraries Australia Help Desk to record enquiries was replaced with RefTracker. The version installed required a form to be used at all times, not a direct email address. This changeover is reflected in the following statistics covering October 2010 to February 2011.

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>telephone</td>
<td>697</td>
</tr>
<tr>
<td>web form</td>
<td>2,797</td>
</tr>
<tr>
<td>email</td>
<td>26</td>
</tr>
<tr>
<td>facsimile</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>3,533</td>
</tr>
</tbody>
</table>

The review of the RefTracker software which supports all Help Desk enquiries found that this product had features comparable with other available products, and while the team will continue to scan offerings in the future, it provides sufficient support for contact with Libraries Australia members.

In March 2011, the opening hours of the Help Desk telephone service were reduced slightly. The Customer Services team is available from 9am until 5pm Australian Eastern Standard time each business day.

Training courses

An Expression of Interest was issued in February 2011 for training providers to give courses in Western Australia, which has been without training services for some
months. Lynn Farkas Information Services Pty Ltd was selected to provide the full suite of Libraries Australia courses on a one-off basis, prior to the Libraries Australia Customer Services team reviewing the offerings in every state and territory.

<table>
<thead>
<tr>
<th>Course</th>
<th># courses</th>
<th>#trainees</th>
<th>Venue</th>
</tr>
</thead>
<tbody>
<tr>
<td>LADD</td>
<td>6</td>
<td>39</td>
<td>South Australia, ACT, Queensland</td>
</tr>
<tr>
<td>LADD refresher</td>
<td>1</td>
<td>1</td>
<td>Queensland</td>
</tr>
<tr>
<td>LA Search</td>
<td>2</td>
<td>23</td>
<td>Queensland</td>
</tr>
<tr>
<td>Libraries Australia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cataloguing Client</td>
<td>1</td>
<td>5</td>
<td>NSW</td>
</tr>
</tbody>
</table>

**Outreach activity**

Libraries Australia staff managed the National Library’s stand at the Information Online conference in Sydney 1-3 February 2011, and gave a presentation on the myths associated with using Libraries Australia. Attendance at an international government libraries Symposium *Fundamental to Democracy*, held at Parliament House from 1-4 March, provided an opportunity to explain the Recent Australian Government publications offering from Libraries Australia², legal deposit obligations and discovery options.

The October and December 2010 eNewsletters were prepared by guest editor and graduate, Sarah Schindeler. The February 2011 issue featured the work of Marian Bate, in particular the Australasian Digital Theses program. An archive is maintained at [www.nla.gov.au/librariesaustralia/news/newsletter.html](http://www.nla.gov.au/librariesaustralia/news/newsletter.html).

Libraries Australia staff attended several User Group meetings in Brisbane, Canberra and Sydney.

The Customer Services team created two new flyers³ and a bookmark explaining the importance of keeping individual entries current in the Australian Libraries Gateway and the InterLibrary Resource Sharing Directory:

---


³ See Attachment A.
Libraries Australia Advisory Committee

The Libraries Australia Advisory Committee (LAAC) expressed its gratitude at the October 2009 meeting to Ms Karen Hansen of the Commonwealth Department of Families, Housing, Community Services and Indigenous Affairs for her participation during 2010 as a special libraries representative. Ms Rosa Serratore, Chief Librarian of the National Meteorological Library at the Bureau of Meteorology, was nominated to replace her, effective December 2010.

Recommendation

The Committee to note the report.

Contact: Debbie Campbell
Director Collaborative Services
dcampbel@nla.gov.au

Contact: Rob Walls
Director Database Services
rwalls@nla.gov.au

16 March 2011
The following libraries began using the Record Import Service during the period 1 October 2010 to 28 February 2011:

- Australian Institute of Music [NAUIM]
- Cairns Regional Council - Cairns City Library [QCCL]
- Charles Sturt University: Regional Archives [NCSRA]
- Department of Environment, Climate Change and Water (NSW) [NEPA]
- Department of Environment, Climate Change and Water (NSW) [NNPW]
- Society of Australian Genealogists [NSAG]
- United Theological College: Camden Theological Library - The Uniting Church in Australia [NUT]
- VicRoads: Business Information Centre [VROAD]
- Whitehorse Manningham Regional Library Corporation: Box Hill Library & Administrative Headquarters [VWMR].
Attachment C

The following libraries joined the Libraries Australia Document Delivery service during the period 1 October 2010 to 28 February 2011:

- BSES Limited [QSB]
- Civil Aviation Safety Authority [ACASA]
- Department of Resources, Energy and Tourism [ARET]
- High Court of Australia Library [AHC]
- Mount Gambier Public Library [SMG]
- NSW Crown Solicitors Office [NCS]
- Queensland Audit Office [QAO]
- Rio Tinto Shared Services Library [QCAL]
- Somerset Regional Council [QSRCL]
- Victorian Department of Transport [VDOI]
- WA Department of Sport and Recreation [WDSR].

In the same period the following libraries withdrew from Libraries Australia Document Delivery:

- ACT Department of Education and Training [AOCIRC]
- Fire & Emergency Services Authority of WA [WFESA]
- TAFE SA Croyden Campus [SFED:CR]
- TAFE SA Adelaide City Campus [SFED:AD]
- WA Department of Consumer and Employment Protection Labour Relations Library [WLPR].
### Libraries Australia Advisory Committee Paper

**LAAC/2011/1/03**


**Report for Libraries Australia Advisory Committee**

**July 2010 to February 2011**

<table>
<thead>
<tr>
<th>Category</th>
<th>Result</th>
<th>Target</th>
<th>% of Target</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Revenue</strong></td>
<td>$2,711,236</td>
<td>$2,689,000</td>
<td>101%</td>
</tr>
<tr>
<td><strong>Expenses</strong></td>
<td>$2,990,621</td>
<td>$3,202,828</td>
<td>93%</td>
</tr>
<tr>
<td><strong>Search Activity</strong></td>
<td>9,768,699</td>
<td>9,340,000</td>
<td>105%</td>
</tr>
<tr>
<td><strong>Holdings Added</strong></td>
<td>2,526,534</td>
<td>1,355,000</td>
<td>186%</td>
</tr>
<tr>
<td><strong>Bibliographic Records Added</strong></td>
<td>702,007</td>
<td>625,000</td>
<td>112%</td>
</tr>
<tr>
<td><strong>Libraries Australia Doc Del Requests</strong></td>
<td>193,096</td>
<td>202,000</td>
<td>96%</td>
</tr>
</tbody>
</table>

**Full Year**

<table>
<thead>
<tr>
<th>Category</th>
<th>Projected</th>
<th>Target</th>
<th>% of Target</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Revenue</strong></td>
<td>$4,042,000</td>
<td>$4,042,000</td>
<td>101%</td>
</tr>
<tr>
<td><strong>Expenses</strong></td>
<td>$4,840,000</td>
<td>$4,840,000</td>
<td>93%</td>
</tr>
<tr>
<td><strong>Search Activity</strong></td>
<td>14,915,000</td>
<td>14,500,000</td>
<td>105%</td>
</tr>
<tr>
<td><strong>Holdings Added</strong></td>
<td>3,550,000</td>
<td>2,100,000</td>
<td>186%</td>
</tr>
<tr>
<td><strong>Bibliographic Records Added</strong></td>
<td>1,100,000</td>
<td>950,000</td>
<td>112%</td>
</tr>
<tr>
<td><strong>Libraries Australia Doc Del Requests</strong></td>
<td>300,000</td>
<td>320,000</td>
<td>96%</td>
</tr>
</tbody>
</table>

**Notes:**

- Includes depreciation, overheads, IT Staff & hardware maintenance costs

---

#### Libraries Australia Revenue and Expenses

![Graph showing Libraries Australia Revenue and Expenses from Jul to Jun]

#### Search Activity

![Bar chart showing search activity from Jul to Jun](chart)

#### Holdings added

![Bar chart showing holdings added from Jul to Jun](chart)

#### Libraries Australia document delivery requests

![Bar chart showing document delivery requests from Jul to Jun](chart)
SUBSCRIPTION MODEL CHANGES FOR 2011

Introduction

Since 2007, the Libraries Australia subscription model has been based on a parameter reflecting the “intrinsic size” of the member library. The parameter varies between sectors. Libraries Australia has already applied such a model to the university library sector (using total library budget as the measure of size), to the state, territory and public library sector (using population served as the measure of size), and to special libraries (offering a choice between total budget and full-time equivalent staff numbers).

During the first quarter of 2011, the Libraries Australia team has reviewed the subscription models for state, territory and public libraries, and made preparations to move approximately 30 libraries not previously migrated to the new special libraries sector model on 1 July 2011. They include tertiary college libraries, and health associations, and the Libraries Australia team is in the process of obtaining budget and staffing data from these libraries.

In 2008, Libraries Australia introduced a new subscription model for the public library and the state/territory library sectors. The model was based on the population figures of each state and territory published by the Australian Bureau of Statistics, included a transition to a fee of 4 cents per head of population to be paid by each public library, and 1.5 cents per head of population to be paid by each state or territory library. It was accompanied by an undertaking to review the model after its first three years.

The reviewed model for State, Territory and public libraries

Only Victoria and the Australian Capital Territory (ACT) have fully transitioned to their endpoints in the population model. Others are still having their annual subscriptions adjusted up or down according to their population figures in 2007. Accordingly, it is proposed to adjust the transition end points and pathways taking into account the most recently published population data. It is proposed that Victoria and the ACT will continue to pay the 4 cent and 1.5 cent fees based on the new population figures.

---

1 New subscription model for state, territory and public libraries from 2008-09 [LAAC/2008/1/4]

2 3218.0 – Regional Population Growth, Australia, 2008-09
The Libraries Australia office anticipates applying the following fee schedules from 1 July 2011.

**State/Territory libraries**

The most recently published population figures will result in the following changes to endpoint subscription fees:

<table>
<thead>
<tr>
<th></th>
<th>Population June 2007</th>
<th>Subscription Endpoint</th>
<th>Population June 2009</th>
<th>Subscription Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSW</td>
<td>6,889,100</td>
<td>$103,337</td>
<td>7,134,421</td>
<td>$107,016</td>
</tr>
<tr>
<td>Victoria</td>
<td>5,205,200</td>
<td>$78,078</td>
<td>5,443,228</td>
<td>$81,648</td>
</tr>
<tr>
<td>Queensland</td>
<td>4,182,100</td>
<td>$62,732</td>
<td>4,425,103</td>
<td>$66,377</td>
</tr>
<tr>
<td>Western Australia</td>
<td>2,105,800</td>
<td>$31,587</td>
<td>2,245,057</td>
<td>$33,676</td>
</tr>
<tr>
<td>South Australia</td>
<td>1,584,500</td>
<td>$23,768</td>
<td>1,625,590</td>
<td>$24,354</td>
</tr>
<tr>
<td>Tasmania</td>
<td>493,300</td>
<td>$7,400</td>
<td>503,292</td>
<td>$7,549</td>
</tr>
<tr>
<td>ACT</td>
<td>339,900</td>
<td>$5,099</td>
<td>352,200</td>
<td>$5,283</td>
</tr>
<tr>
<td>Northern Territory</td>
<td>215,000</td>
<td>$3,225</td>
<td>225,938</td>
<td>$3,389</td>
</tr>
<tr>
<td>Total</td>
<td>21,014,900</td>
<td>$315,224</td>
<td>21,952,829</td>
<td>$329,292</td>
</tr>
</tbody>
</table>

Based on the new subscription endpoints, the following fees will apply to the State and Territory libraries:

<table>
<thead>
<tr>
<th></th>
<th>2010-11</th>
<th>2011-12</th>
<th>2012-13</th>
<th>2013-14</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSW</td>
<td>$132,000</td>
<td>$129,451</td>
<td>$122,979</td>
<td>$116,830</td>
</tr>
<tr>
<td>Victoria</td>
<td>$78,078</td>
<td>$81,648</td>
<td>$81,648</td>
<td>$81,648</td>
</tr>
<tr>
<td>Queensland</td>
<td>$35,000</td>
<td>$40,208</td>
<td>$44,229</td>
<td>$48,652</td>
</tr>
<tr>
<td>Western Australia</td>
<td>$53,000</td>
<td>$52,240</td>
<td>$48,240</td>
<td>$44,240</td>
</tr>
<tr>
<td>South Australia</td>
<td>$34,000</td>
<td>$31,765</td>
<td>$28,765</td>
<td>$25,765</td>
</tr>
<tr>
<td>Tasmania</td>
<td>$33,000</td>
<td>$29,587</td>
<td>$25,587</td>
<td>$21,587</td>
</tr>
<tr>
<td>ACT</td>
<td>$5,099</td>
<td>$5,283</td>
<td>$5,283</td>
<td>$5,283</td>
</tr>
<tr>
<td>Northern Territory</td>
<td>$27,000</td>
<td>$23,119</td>
<td>$19,119</td>
<td>$15,119</td>
</tr>
<tr>
<td>Total</td>
<td>$397,177</td>
<td>$393,302</td>
<td>$375,851</td>
<td>$359,125</td>
</tr>
</tbody>
</table>

It should be noted that in Western Australia, Tasmania, the ACT and the Northern Territory, the respective State and Territory libraries pay the subscription fees on behalf of their public libraries.

**Public libraries**

The most recently published population figures will result in the following changes to endpoint subscription fees:
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NSW</td>
<td>6,889,100</td>
<td>$275,564</td>
<td>7,134,421</td>
<td>$285,377</td>
</tr>
<tr>
<td>Victoria</td>
<td>5,205,200</td>
<td>$208,208</td>
<td>5,443,228</td>
<td>$217,729</td>
</tr>
<tr>
<td>Queensland</td>
<td>3,720,454</td>
<td>$148,818</td>
<td>4,052,106</td>
<td>$162,084</td>
</tr>
<tr>
<td>Western Australia</td>
<td>2,105,800</td>
<td>$84,232</td>
<td>2,245,057</td>
<td>$89,802</td>
</tr>
<tr>
<td>South Australia</td>
<td>1,584,500</td>
<td>$63,380</td>
<td>1,623,590</td>
<td>$64,944</td>
</tr>
<tr>
<td>Tasmania</td>
<td>493,300</td>
<td>$19,732</td>
<td>503,292</td>
<td>$20,132</td>
</tr>
<tr>
<td>ACT</td>
<td>339,900</td>
<td>$13,596</td>
<td>352,200</td>
<td>$14,088</td>
</tr>
<tr>
<td>Northern Territory</td>
<td>215,000</td>
<td>$8,600</td>
<td>225,938</td>
<td>$9,038</td>
</tr>
<tr>
<td></td>
<td>20,553,254</td>
<td>$822,130</td>
<td>21,579,832</td>
<td>$863,193</td>
</tr>
</tbody>
</table>

The total population figure is lower (than the State and Territory table) as not all Queensland public libraries are Libraries Australia members.

Based on the new subscription endpoints, the following fees will apply to public libraries:

<table>
<thead>
<tr>
<th></th>
<th>2010-11</th>
<th>2011-12</th>
<th>2012-13</th>
<th>2013-14</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSW</td>
<td>$335,945</td>
<td>$335,744</td>
<td>$318,957</td>
<td>$305,619</td>
</tr>
<tr>
<td>Victoria</td>
<td>$206,263</td>
<td>$217,729</td>
<td>$217,729</td>
<td>$217,729</td>
</tr>
<tr>
<td>Queensland</td>
<td>$147,045</td>
<td>$158,297</td>
<td>$162,084</td>
<td>$162,084</td>
</tr>
<tr>
<td>Western Australia</td>
<td>$37,000</td>
<td>$49,042</td>
<td>$58,042</td>
<td>$67,042</td>
</tr>
<tr>
<td>South Australia</td>
<td>$38,000</td>
<td>$42,011</td>
<td>$45,011</td>
<td>$48,011</td>
</tr>
<tr>
<td>Tasmania</td>
<td>$6,000</td>
<td>$8,162</td>
<td>$10,162</td>
<td>$12,162</td>
</tr>
<tr>
<td>ACT</td>
<td>$13,596</td>
<td>$14,088</td>
<td>$14,088</td>
<td>$14,088</td>
</tr>
<tr>
<td>Northern Territory</td>
<td>$5,000</td>
<td>$7,356</td>
<td>$8,756</td>
<td>$9,038</td>
</tr>
<tr>
<td></td>
<td>$788,850</td>
<td>$832,430</td>
<td>$834,830</td>
<td>$835,774</td>
</tr>
</tbody>
</table>

The next review for these sectors will be scheduled in 2014.

**Recommendation**

The Libraries Australia Advisory Committee to approve the proposed changes.

**Contact:**

Warwick Cathro  
Assistant Director-General  
Resource Sharing & Innovation Division  
wcathro@nla.gov.au  

Debbie Campbell  
Director Collaborative Services  
dcampbel@nla.gov.au  

17 March 2011
Changes to the Libraries Australia User Agreement 2011

Introduction

Each year, the Libraries Australia User Agreements are reviewed to ensure their currency, particularly for new members. The scope of the changes is generally modest.

This year, several changes have been proposed to clarify the obligations and benefits of membership. The changes do not affect the intent of previously signed versions of the User Agreement. They have been applied to all three versions of the Agreement – for libraries, for commercial providers, and for individuals (note that the wording quoted below cites clause numbers for the libraries’ version).

As part of this review, Product pricing information has also been updated. Refer to Appendix A of this paper for further discussion.

Additions to the User Agreement

1. At Clause 3.5, a new sentence to highlight the nature of the cooperative:

   “Libraries Australia Services are sustained through national cooperation and rely on the practice of reciprocity in all dealings between member libraries, including the User. As such, members should ensure that they will not engage in activities that diminish the value of Libraries Australia. In particular, the User must not:”

   (The remainder of the wording in this paragraph is unchanged.)

2. At Clause 8.1 in Attachment A, a reminder of the option to be a member of WorldCat:

   “8. Membership of WorldCat

   8.1 Through membership of the Libraries Australia, as evidenced by this Agreement, the User is entitled:
   (a) to membership of WorldCat, provided the User contributes bibliographic and/or holdings records to WorldCat;
   (b) as a member, to nominate themselves or others, and vote for, a national representative on the OCLC Asia Pacific Regional Council.

   Further information regarding OCLC governance arrangements is summarised at <www.nla.gov.au/librariesaustralia/ocl.html>.“
3. At Clause 4.3.4 in Attachment B, a new charge for Product orders from consortia:

“4.3.4 Product charges for consortia

When a consortium of libraries requests a Products file, a set-up charge of $AUD200 will apply. Record charges for files containing more than 2,000 will be set at $AUD0.66 per record, and the concessional rate in clause 4.3.2 will also apply.”

Changes to the User Agreement

1. In Attachment A, clause 2.1, the reduction of telephone support to the hours of 9am-5pm AEST, mainly due to the increase in queries via the online contact form.

2. Most Product pricing in Attachment B of the User Agreement has been updated.

3. The most significant change to charging is at Attachment B, clause 4.2: “A request to provide a one-off Product within a 48-hour turnaround time will attract an additional charge of $550.”

Deletions from the User Agreement

1. Due to the merger of the Australian Research Online, Picture Australia and Music Australia services into Trove, references to these services have been removed;

2. Due to the inclusion of APAIS and AMI metadata in Trove, references to APAIS searching charges have been removed.

Recommendation

The Libraries Australia Advisory Committee to endorse the changes.

Contact: 
Debbie Campbell
Director Collaborative Services
dcampbel@nla.gov.au

17 March 2011
Appendix A

Since 2007, the Libraries Australia team has been reviewing and renewing subscription models for all Libraries Australia services. However, the charges for Libraries Australia Products have not changed since the service went live in 2005/06¹.

Issues

During the last three years, the use of the Products suite has declined significantly. In 2010, the suite was promoted in the Libraries Australia newsletter and with a bespoke advertisement in InCite magazine, with no discernible impact.

Overview

The total revenue raised from the Products suite has been:

<table>
<thead>
<tr>
<th>Year</th>
<th>Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007/08</td>
<td>$28,397</td>
</tr>
<tr>
<td>2008/09</td>
<td>$31,311*</td>
</tr>
<tr>
<td>2009/10</td>
<td>$19,996</td>
</tr>
<tr>
<td>2010/11</td>
<td>$14,000**</td>
</tr>
</tbody>
</table>

Table 1: total and projected revenue in recent years

*includes special requests for collection analysis projects at the University of New South Wales

** projected end of year budget is $25,000-$30,000.

The total number of records provided in 2009/10 was 49,323,455; down from 69,549,086, although the latter was primarily due to the large number of records extracted for collection analysis purposes. The following table shows the total number of products produced during the last three financial years.

<table>
<thead>
<tr>
<th>Job type</th>
<th>Number of runs (2007/08)</th>
<th>Number of runs (2008/09)</th>
<th>Number of runs (2009/10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charged for</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extractions for vendors</td>
<td>90</td>
<td>76</td>
<td>69</td>
</tr>
<tr>
<td>Regular subscription files</td>
<td>122</td>
<td>146</td>
<td>113</td>
</tr>
<tr>
<td>More than 2,000 records / file</td>
<td>75</td>
<td>65</td>
<td>11</td>
</tr>
<tr>
<td>Not charged for</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintenance</td>
<td>455</td>
<td>356</td>
<td>366</td>
</tr>
<tr>
<td>Acquisitions</td>
<td>546</td>
<td>564</td>
<td>489</td>
</tr>
<tr>
<td>Less than 2,000 records / file</td>
<td>634</td>
<td>387</td>
<td>362</td>
</tr>
<tr>
<td>Total</td>
<td>1,922</td>
<td>1,594</td>
<td>1,410</td>
</tr>
</tbody>
</table>

Table 2: total product runs

Notes

1. Acquisition files are produced for the NSLA libraries (currently excluding South Australia and Tasmania). Each requires a sophisticated query against

the NBD to meet complex acquisition policies. They are run quarterly and are provided free of charge.

2. Almost all maintenance runs are for internal purposes only, and are not eligible for revenue raising. Some of the files support our international collaboration with NACO (the Name Authority Cooperative) and the VIAF (Virtual International Authority File) project, both US-based.

3. Member libraries are charged for orders resulting in files greater than 2,000 records. The total number of Products supplied externally in recent years was as follows:

4. Regular subscription files contain ‘e-collection’ sets or authority record files. They are billed annually but provided quarterly.

5. Extractions for vendors attract a different per record charge.

6. The Libraries Australia team will continue to provide one-off quotes for special requests which take into account the amount of support given by staff.

Further considerations

Human intervention is required to configure regular product jobs and run one-off enquiries. Infrequent large file extractions impact on search database performance and are therefore scheduled outside business hours.

---

Only 29 Libraries Australia members regularly request products. In 2009/10, only 11 files containing more than 2,000 records were supplied; while 298 files contained less than 1,000 records. (The remainder contained between 1,000 and 2,000 records).

It is not proposed to increase the cost of providing LADD Location data at this time. An investigation to determine whether more files, such as Title Lists, could be made self-serve in a method similar to Recent Australian Publications (RAP) and Recent Australian Government Publications (GovRAP) will be scheduled within the Libraries Australia Search platform replacement project.
TROVE STATUS REPORT

Introduction

Trove, the National Library’s premier national discovery service, has now been in operation for more than a year. Since the last report to the Libraries Australia Advisory Committee in October 2010, the following enhancements have been implemented:

- the integration of functions from the separate digitised Australian newspaper service;
- the integration of the Australasian Digital Theses program (ADT) and the Australian Research Online service;
- the addition of the Australian Women’s Weekly from 1933 to 1982 as digitised articles in the Digitised newspapers and more zone;
- two new releases of Trove software which fixed bugs and implemented user enhancements (v3.02 on 10 December and v3.03 on 25 January). Users are now able to be alerted to new content with keywords of interest and newspaper article alerting is improved. In addition work was undertaken in conjunction with the ALG to make library access visible and to enable searching by library in the advanced search screens; and
- a method to expedite the addition of collection lists described in MARC and submitted to Trove via the National Bibliographic Database. The lists are important for bringing awareness to unique and special collections.

In addition, preparations for the integration of the separate Picture Australia and Music Australia services have commenced and an Application Programming Interface (API) was scoped. To assist with the latter, a requirements survey was circulated and resulted in significant interest from a diverse range of service providers.

New contributors which have not participated in any National Library services before were invited to join Trove, and several significant collections are being mapped and tested, including from the National Film and Sound Archive, the Australian Broadcasting Commission, as well as the Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS).

The amount of new content in Trove has increased significantly as follows:

<table>
<thead>
<tr>
<th>Zone Name</th>
<th>Work Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>All view (as at February 2011)</td>
<td>120,102,080</td>
</tr>
<tr>
<td>Australian Newspapers</td>
<td>45,313,730</td>
</tr>
</tbody>
</table>
### Stage 4 development

Authorised access to licensed eJournals is in development as the fourth and final stage of the Trove project. The National, State and Territory Libraries (NSLA group) has assisted via its Reimagining Libraries project Open Borders, by providing details about different authentication options such as EZproxy servers and recognition of barcodes printed on library cards. Six university libraries also participated by including their OpenURL resolvers.

The Australian Libraries Gateway was expanded to store information about and support the maintenance of the authentication mechanisms.

Two vendors, Cengage Gale and RMIT Publishing, have worked with the National Library to make more than 40 million articles accessible via Trove. As this is expected to increase during 2011, interface design for a new zone has been completed and changes will be made to the Trove homepage to reflect this.

### Usage

By February, the number of registered users in Trove was 27,500.

The last six months have also seen the Trove Communications and Marketing Plan come to fruition, with a particular emphasis on growing the social engagement response to Trove through Twitter and the blog in the Trove User Forum at http://trove.nla.gov.au/forum/. By March, the number of Twitter followers registered with @TroveAustralia was 500 and included several media outlets. The Trove team has set up a roster to add entries to the Twitter feed and the Trove blog on a daily basis. A Facebook account has been less successful.

Several members of the public have their own blogs which feature Trove content, including:

- [Now and then Snippets of Australian social history from newspapers and other sources](http://lynnwalsh.wordpress.com/)
- [Historian at work](http://historianatwork.wordpress.com/2011/01/07/a-dose-of-troveaustralia-love/), and

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Books, journals, magazines, articles</td>
<td>31,347,699</td>
</tr>
<tr>
<td>Pictures and photos</td>
<td>4,266,993</td>
</tr>
<tr>
<td>Archived websites</td>
<td>35,981,716</td>
</tr>
<tr>
<td>Music sound and video</td>
<td>1,453,639</td>
</tr>
<tr>
<td>People and organisations</td>
<td>882,820</td>
</tr>
<tr>
<td>Diaries, letters, archives</td>
<td>513,361</td>
</tr>
<tr>
<td>Maps</td>
<td>342,122</td>
</tr>
</tbody>
</table>
Searching activity in Trove continues to increase steadily, with more than 30,000 visits per day, reaching almost three million unique visitors during 2010. They executed 22 million searches between February and December 2010. This compares to 2.7 million searches of Libraries Australia during 2009, when it offered the equivalent free searching of Libraries Australia. This target will be surpassed in 2011, as more than 12 million searches, have occurred in January and February alone. This is partly due to the integration of the digitised newspaper content. The number of pageviews, which illustrate the stickiness of the site, reached 126,715,445.

The number and types of enquiries received were:

<table>
<thead>
<tr>
<th>Enquiry type</th>
<th>Nov # enquiries</th>
<th>Dec/Jan # enquiries</th>
<th>Feb # enquiries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change Trove records</td>
<td>7</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Digitised newspapers and more</td>
<td>75</td>
<td>103</td>
<td>73</td>
</tr>
<tr>
<td>Get a Trove item</td>
<td>6</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>Trove feedback</td>
<td>3</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>Trove functionality</td>
<td>207</td>
<td>52</td>
<td>116</td>
</tr>
<tr>
<td>Trove technical problem</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Trove username</td>
<td>0</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>299</strong></td>
<td><strong>188</strong></td>
<td><strong>222</strong></td>
</tr>
</tbody>
</table>

Other types of use during 2010 were:

<table>
<thead>
<tr>
<th>Trove user engagement</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lines of text corrected in Newspapers</td>
<td>26,259,688</td>
</tr>
<tr>
<td>Tags added</td>
<td>685,258</td>
</tr>
<tr>
<td>Images added by the public via Flickr</td>
<td>88,951</td>
</tr>
<tr>
<td>Users registered</td>
<td>24,619</td>
</tr>
<tr>
<td>Comments added</td>
<td>16,040</td>
</tr>
<tr>
<td>Records Merged</td>
<td>9,297</td>
</tr>
<tr>
<td>Lists created</td>
<td>2,090</td>
</tr>
<tr>
<td>Records split</td>
<td>1,936</td>
</tr>
<tr>
<td>Messages in User Forum</td>
<td>546</td>
</tr>
<tr>
<td>Use of mobile devices (iPhone, iPad, iPod)</td>
<td>49,218</td>
</tr>
</tbody>
</table>
Impact

The 2011 floods in Queensland generated a lot of interest in details about comparative floods, particularly in 1973 and 1894. The Rockhampton Heritage Collections Officer at the Rockhampton Regional Council Library advised the Trove team in January that a small group of volunteers known as the Treasure Trovers were aiming to ensure that Rockhampton newspapers would “be the first newspaper on the Trove site that’s as near to 100% corrected as possible. I’ve been surprised how invested and how much ownership the group has taken of our paper - we are meeting even in the middle of the flood crisis!”

As mentioned in the October report¹, Trove has also been chosen by several strategic academic projects as a research tool, centred on the newspaper content. Its integration into Trove in November 2010 resulted in additional media coverage:

1 Trove report 2010

¹ Trove report 2010
Volunteers with an eagle-eye on the news
Conrad Walters
February 7, 2011

Leading the way ... Rose Holley makes old newspapers available online. Photo: Vikky Wilkes
ANN MANLEY is addicted to correcting types.

Manley is among thousands of volunteers who correct transcripts of old newspapers for the National Library of Australia. The practice of recruiting strangers for such work is called crowd sourcing and the library has quietly built an international reputation as a pioneer in it.

The idea has paid off. In the past few weeks the library recorded the biggest-ever use of its online newspaper archives as people sought reports on Brisbane's 1974 floods.

Another big contributor, Julie Hempstead, of Victoria, recently corrected an article from 1890 about Ned Kelly's murder trial.

Such volunteers are the mainstay of the library's goal of making the nation's old newspapers available online and searchable, said Trove's manager, Rose Holley.


The Trove team continued to receive extraordinary feedback via the Contact form, and a recent example was as follows:

“I have just discovered your site...having googled the name of my great grandfather who was a well known Queenslander in the late 1800s/early 1900s I have come across hundreds of references to him, letters written by him to newspapers and all sorts of things.
I knew something about him but was lacking much of this minutiae, I am absolutely transfixed by what you have provided and suddenly feel that I know him so much better.
Thank you! This is a brilliant concept and very well executed.”

Resources

A video describing Trove was prepared as an introduction to the service, and is highly recommended for use by the schools sector. It is complemented by several screencasts which focus on specific functions of the service such as how to register for a personal account, creating a list of selected items, tagging and correcting newspaper text. They are available at:

- general introduction <www.youtube.com/watch?v=DDOY5gun1mY> and
- screencasts <http://www.youtube.com/user/TroveNLA>.
Some supporting documents have also been made available:

- authorised access to licensed journals (Stage 4 of Trove)  
- Adding context to resources – the ‘List’ feature in Trove September 2010  
- Ideas for how Volunteers at Cultural Heritage Institutions can help, using Trove as a Tool  
- How to arrange for finding aids, transcripts and lists to be indexed into Trove  

The latter document was a product of the work of the Reimagining Libraries’ Description and Cataloguing project2.

Marketing

Trove marketing items have included a notebook and a USB flash drive, distributed to professional conferences in the library sector and family history societies; posters and bookmarks distributed to 11,000 schools and music libraries; and more modest numbers of T-shirts and mousemats which are available for purchase in the Library’s online bookshop.

Paper-based items will be updated with a new Trove logo which is currently being registered with IP Australia, and new promotional items will be considered.

A detailed report of Trove’s first year, including presentations, was prepared by the Trove Manager - <www.nla.gov.au/openpublish/index.php/nlasp/article/view/1882>.

Professional contact

Trove has also inspired other agencies such as *The New York Times* to approach the Trove team for advice on the text correction concept, or the National Library of Finland to replicate the functionality.

**National Library of Finland Turns to Crowdsourcing, Games to Help Digitize Its Archives**  
**By Audrey Watters (February 8, 2011 8:02 AM / 0 Comments**

The National Library of Finland has launched a new program to support the digitization efforts of its archives. The project, Digitoolkot (Digital volunteers), blends microtasks, crowdsourcing, and video games to break up and distribute some of the dull repetitive work of verifying digitized records.

*"We have millions and millions of pages of historically and culturally valuable documents, newspapers and journals online. The challenge is that the optical character recognition often contains errors and omissions, which hang even for example searches," says Kai Elholm, Director of the National Library of Finland. "Manual correction is needed to weed out these mistakes so that the texts become machine readable, enabling scholars and archivists to search the material for the information they need."

Acknowledgements

Rose Holley, Kaye McIntyre, Alison Dellit, Virginia James, Mark Raadgever, Julia Hickie, Mary-Louise Weight, Sarah Schindeler and the Trove IT team brought passion, diligence, and unflagging enthusiasm to this phase of Trove development.

Recommendation

The Libraries Australia Advisory Committee to note the report.

Contact:  
Debbie Campbell  
Director Collaborative Services  
dcampbel@nla.gov.au

16 March 2011
Revision of the Resource Sharing and Innovation Strategic Plan

Introduction

The Strategic Plan has been revised to reflect activities that have been completed, new activities that have commenced, and changes in priorities since it was prepared in mid-2010.

The next full revision of the Plan will be undertaken in mid-2011. The Advisory Committee is invited to provide input for the forthcoming revision.

The following is a summary of the changes to the current Plan.

The revised plan is attached.

Additions to the Strategic Plan

The following activities have been added:

- Integrate the separate discovery services into Trove;
- Review the options for the continuing operation and development of LADD including the strategic direction of the VDX software and the future of the ISO ILL protocol;
- Communicate with member libraries through sectoral events;
- Implement system interoperability and contractual arrangements to enable LADD users to forward requests to British Library Document Supply Centre (BLDSC).

Deletions from the Strategic Plan

The following activities have been deleted:

- Renewing the agreement with OCLC which was signed in 2007;
- Creating an online User Forum for Trove to enable social engagement and enhancements to data engagement functions including lists and tags;
- Adding support to Trove to deliver the Australian Womens Weekly and other journal content.

Rewording to more accurately describe the activity

The descriptions of the following activities have been revised:
• Integrate the Australian Libraries Gateway and the Inter Library Resource Sharing Directory, and explore the feasibility of automated migration of the data into the WorldCat Registry;

• Work with member libraries, OCLC, and if necessary system vendors to encourage the implementation of the SRU Record Update protocol, or the WorldCat Synchronisation widget, so that local systems of member libraries can update the ANBD as a real-time background task.

**Recommendation**

The Libraries Australia Advisory Committee to note the changes.

The Libraries Australia Advisory Committee to provide input for the forthcoming revision of the Strategic Plan.

Contact: Warwick Cathro  
Assistant Director-General  
Resource Sharing and Innovation  
wcathro@nla.gov.au

28 March 2011
STRATEGIC PLAN

JULY 2010 TO JUNE 2012

Revised
February 2011
## CONTENTS

1. EXECUTIVE SUMMARY .......................................................... 3

2. INTRODUCTION ................................................................. 5

3. ENVIRONMENTAL SCAN ...................................................... 6
   User expectations and characteristics ..................................... 6
   Changes in libraries and other collecting institutions ...................... 6
   Other environmental issues ................................................... 7
   Discussion ............................................................................. 7

4. LIBRARIES AUSTRALIA ....................................................... 8
   Service overview ................................................................. 8
   Value Proposition ................................................................... 9
   Underlying system ............................................................... 9
   Strategies for July 2010 to June 2012 ...................................... 10

5. E-RESOURCE CONSORTIA ..................................................... 15
   Service overview ................................................................. 15
   NSLA E-Resources Consortium ............................................... 16
   Electronic Resources Australia (ERA) ....................................... 16
   Licence conditions .................................................................. 17
   Strategies for July 2010 to June 2012 ...................................... 18

6. DISCOVERY SERVICES FOR THE PUBLIC ............................ 19
   Overview of existing services ............................................... 19
   Strategies for 2009-2011 ....................................................... 20

7. THE WIDER COLLABORATIVE AGENDA ............................... 22
   Reimagining Libraries .......................................................... 22
   Party Infrastructure Project ................................................... 22
   Standards activities ............................................................. 23
1. EXECUTIVE SUMMARY

This document sets out the plans of the Resource Sharing and Innovation Division for the development and ongoing delivery of its services between July 2010 and June 2012.

The Division serves four groups of users:

- those Australian libraries that are members of Libraries Australia;
- those Australian libraries that participate in consortial e-resource licensing arrangements that are co-ordinated by the Division;
- the researchers and public of Australia, who discover, access and annotate collection resources through the Library’s flagship discovery service (Trove) and other discovery services which Trove will eventually replace; and
- the libraries, cultural institutions, universities and other organisations with whom the Library collaborates, including those who contribute data for the Division’s services.

These user groups relate to the four “service bundles” that are described in the body of this Plan.

The Plan takes account of changes in user expectations, technology, and the wider environment, including:

- the upsurge in community participation in social networking activities;
- the expectation of users that access to much content will be free;
- ongoing pressure on library budgets, which continues to press libraries to review their activities, increase the efficiency of services and do more with less;
- the adoption by many libraries of integrated discovery tools such as Primo and Summon which are transforming and extending local catalogues; and

For Libraries Australia, planned activities include:

- extending and maintaining partnerships with member libraries, promoting the Libraries Australia value proposition;
- extending the subscriber base, particularly for special and school libraries;
- maintaining the subscription model to ensure an equitable sharing of financial support for the service;
- scoping and commencing a project to redevelop the Libraries Australia Search software module;
- exploring the feasibility of migrating the Australian Libraries Gateway and the Inter Library Resource Sharing Directory into the WorldCat Registry;
- exploring with selected member libraries (particularly state and territory libraries) the feasibility of enabling their local discovery services to become “views” of Trove;
- working systematically with member libraries to eliminate gaps in their data contributions and to encourage methods of more timely contribution of data;
- improving reporting of holdings deletions from public library de-selection activities;
- working with member libraries and OCLC to implement new technical mechanisms for more timely reporting of data to the ANBD;
- continuing to enrich the ANBD by harvesting additional value-added content, such as Table of Contents data from the Library of Congress and other sources;
supporting the implementation of the new cataloguing standard RDA (Resource Description and Access);  
- maintaining strong relationships with OCLC staff in Ohio, Leiden and Melbourne;  
- investigating the feasibility of the interconnection of WorldCat Resource Sharing with LADD; and  
- supporting Australian representation on the Asia Pacific Regional Council of OCLC, and participating in forums organised by the Council.

For the **E-Resource Consortia**, planned activities include:

- promoting the ERA products widely to appropriate libraries;  
- updating the ERA product offerings annually;  
- lobbying for funding to support a core set, either for the whole community or for a specific sector such as schools;  
- negotiating with vendors in an effort to obtain more attractive whole-of-sector prices;  
- obtaining whole-of-NSLA prices for the new NSLA core set, with the objective of taking up those licences wherever possible;  
- streamlining the ERA pricing process, including the abolition of “product polls”; and  
- participating in Project 7 (“Collaborative Collections”) of the NSLA Reimagining Libraries initiative.

For **Trove and the other Discovery Services**, planned activities include:

- actively marketing Trove, and encouraging users to register, so that they can interact with and contribute to the content in Trove;  
- introducing an online User Forum;  
- ensuring that the *Australian Womens Weekly* and other journal content can be delivered through Trove;  
- adding a range of interface and search improvements in response to the outcomes of a usability review;  
- scoping a standards-based programmer API, to allow other discovery services to use Trove as a target;  
- adding journal article content and e-resource metadata, with support for access to the full text of e-resources where that can be easily implemented;  
- engaging with a wide range of collecting institutions, including archives, museums and galleries, universities and other research repositories;  
- adding data from major cultural institutions;  
- establishing an online Trove Contributor forum so that data contributors can communicate more easily with each other and with the Division’s staff; and  
- integrating separate discovery services into Trove.

The Division will also engage in a broad set of **collaborative activities** which support the Library’s Directions and also support the future development of the three service bundles discussed above. These activities include:

- participation in Reimagining Libraries projects;  
- participation in relevant standards activities, such as the Australian Committee on Cataloguing and the development of Encoded Archival Context (EAC); and  
- participation in the Party Infrastructure Project, including extending the People Australia Data Administration module to enable distributed name authority control and providing machine interfaces to data about researchers and research organisations.
2. INTRODUCTION

The National Library Act 1960 identifies four functions for the Library. The third and fourth of these functions are:

- to make available “such other services in relation to library matters and library material (including bibliographical services) as the Council thinks fit” ...
- to cooperate in library matters ....

These are the functions that relate most closely to the work of the Resource Sharing and Innovation Division.

The Library’s Directions for 2009-2011 includes these statements:

- **We will use innovative technologies** ... to streamline our services, providing users with simple and easy access to our collections and access to the world’s knowledge.
- **We have a national focus and are committed to taking a leadership role in the Australian cultural community.**
- **We will collaborate with a variety of other institutions to improve the delivery of information resources to the Australian public.**
- **We will continue to maintain and develop the Libraries Australia service to ensure that it meets the developing needs of its member libraries and their users.**

The Directions for 2009-2011 also identifies these activities:

- **Introduce a new online service incorporating all the information available through our existing discovery services (this will include access to greater full-text content along with facilities for users to comment and add to our data)**
- **Develop Electronic Resources Australia in order to deliver a core set of information resources to the Australian public**
- **Create partnerships with libraries, cultural institutions, universities and government agencies to improve the data available through our new online discovery service.**

These over-arching statements of intent have guided the development of this Strategic Plan.

The Division serves four groups of users:

- those Australian libraries that subscribe to Libraries Australia;
- those Australian libraries that participate in consortial e-resource licensing arrangements that are co-ordinated by the Division;
- the researchers and public of Australia who discover, access and annotate collection resources through the Library’s flagship discovery service (Trove) and the other discovery services which Trove will eventually replace; and
- the libraries, cultural institutions, universities and other organisations with whom the Library collaborates, including those who contribute data for the Division’s services.

These user groups are served by the three service bundles and the wider collaborative agenda that are described in the body of this Plan. For each of the service bundles, the Plan will provide a Service Overview and a set of Strategies for July 2010 to June 2012.
3. ENVIRONMENTAL SCAN

There are a number of trends in user expectations, in the library sector, and elsewhere in the environment which have the potential to influence this Plan.

User expectations and characteristics

Notable trends in user expectations and characteristics include:

- increased end-user demand for easily accessible online resources for their research and information needs, irrespective of the user’s location;
- the widespread practice of users to commence their information seeking at search engines, rather than in library catalogues or portals;
- the expectation of users that access to much content will be free;
- the expectation of users that they will be able to access information resources through their smart mobile devices;
- an increase in the older population, with an associated increase in independent research by the “citizen-scholar”; and
- the upsurge in community participation in social networking activities, both through spaces such as Facebook and through annotation, commenting, tagging and content creation in information delivery spaces.

Changes in libraries and other collecting institutions

Notable trends in the library sector include:

- ongoing pressure on library budgets, which continues to press libraries to review their activities, increase the efficiency of services and do more with less;
- a modest decline in the number of Australian libraries due to public library amalgamations and the abolition of some special libraries;
- declining print collections, especially in university and special libraries;
- a greater emphasis by libraries on their collections of licensed e-resources;
- increasing ‘homogenisation’ of collections due to a preference for more economical ‘bundled’ subscriptions, rather than more expensive subscriptions to individual titles;
- the widespread free availability of metadata, including MARC records, which undermines any business model based largely on charging for provision of such data;
- a realisation by university libraries that their special, unique or ‘hidden’ collections need to be made more accessible;
- the need for special libraries to increase awareness of the national value of their unique and often hidden collections, a finding supported by recent surveys;
- the existence of the Reimagining Libraries initiative of National & State Libraries Australasia;
- the adoption by many libraries of integrated discovery tools such as Primo and Summon which are transforming and extending local catalogues;
- the outsourcing of technical services processes by many libraries, particularly public and school libraries;
- patchy but continuous progress in digitising library collections, and collections of other cultural institutions;
- an interest by many stakeholders in the provision, perhaps by the National Library, of a national digital archive to store and manage digital collection content where no local solution is available;
 the survival of inter-library lending, or at least the slow pace of its decline, despite the fact that most users are strongly discouraged by barriers to ‘getting’ resources;
 the development and maintenance of institutional research repositories in university libraries;
 a growing trend for library systems to become more modular, open and cloud-based; and
 libraries looking more to externally hosted services and outsourcing IT infrastructure support.

Other environmental issues

Other notable environmental issues include:
 ongoing development of search engines, including the proliferation of new Google services;
 the impending global availability of the Google Books database as an e-resource offering;
 the continued expansion of OCLC’s services including WorldCat;
 increased interest in use of “linked data”, the semantic web and Application Programming Interfaces (APIs);
 increased use of the Web to “mash up” data and to repurpose data in other spaces;
 hesitation by some e-resource vendors about making their metadata available to libraries for re-purposing in cross-vendor discovery services; and
 lack of action to date by government in funding national licensing for a core set of e-resources.

Discussion

It follows from the above that:
 Libraries Australia will continue to support national shared cataloguing and inter-library loan services;
 Libraries Australia will need to continue to operate as efficiently and cost effectively as possible, in order to moderate its subscription fees;
 Trove will aim to facilitate end user access to resources that form part of libraries’ e-resource subscriptions;
 Libraries Australia will continue to broker deals which make useful metadata content available to the member libraries (as has happened with WorldCat and Blackwells);
 Libraries Australia will work with university, state, public and special libraries on strategies to make “hidden collections” more accessible;
 Libraries Australia could undertake a project to improve both awareness and accessibility of hidden collections in special libraries;
 the licensing of e-resources should be as affordable and flexible as possible for the libraries that use the Division’s services;
 the library sector should attempt to persuade Government to fund at least some library sectors (eg schools) in licensing a core set of e-resources;
 the Division will need to maintain contact and dialogue with library system vendors;
 the Division should enter into a dialogue with e-resource vendors concerning the release of their metadata for re-purposing in Trove;
 the content of Trove should be exposed comprehensively in search engines;
 Trove will develop an API;
Resource Sharing & Innovation

Strategic Plan

Last updated 13/04/2011

Page 8 of 24

- Trove will aim to include digitised collection content wherever possible;
- the Library should strive to make metadata freely available in Trove wherever possible;
- Trove will encourage its users to exploit its annotation, tagging and content creation facilities; and
- Trove will need to develop interfaces for smart mobile devices.

4. LIBRARIES AUSTRALIA

Service overview

Australian libraries have traditionally placed a high value on collaboration and resource sharing. In this context, Libraries Australia and its predecessor services, based on the Australian National Bibliographic Database (ANBD), have been provided by the Library since 1981. The service is a fundamental part of Australia’s library and information infrastructure.

The purpose of Libraries Australia is to support the collection management workflows of its member libraries. It supports collaboration between those members, including cooperative cataloguing, enabling the member libraries to reduce the costs of their cataloguing by using records created by other libraries in Australia and worldwide. In order to improve access to library collections, Libraries Australia supports interlending and efficiently manages the associated payments.

Libraries Australia serves over 1,200 Australian libraries including all state, university and most special, government and public libraries.

Libraries Australia’s key asset is its underlying database (the Australian National Bibliographic Database, or ANBD) containing records of over 47 million items (in December 2010) held by about 700 member libraries. Subscribers can also access major international databases including WorldCat (provided by OCLC) and Te Puna (New Zealand libraries’ national union catalogue).

Libraries Australia and its predecessor services have always operated on a partial cost recovery basis. While the National Library has always provided a subsidy, the service has historically operated at or slightly above marginal cost recovery, allowing the Library to support continued growth in the service.

Consultation with member libraries on their needs and the future development of the service occurs through the Libraries Australia Advisory Committee, and through regular meetings with member libraries in all states and territories.

A new subscription model for Libraries Australia was announced in 2006 and progressively implemented between 2007 and 2010. The new model aims to encourage unlimited use by member libraries. It was applied to university libraries in 2007 (and updated in 2010); to state, territory and public libraries in 2008 (updated in 2011); to special and school libraries and commercial users in 2009; and to TAFE libraries in 2010. The implementation of the model, including the subscription tiers, will be reviewed for each sector at three year intervals.
Value Proposition
This Plan assumes that there remains a strong imperative for the provision of the Libraries Australia service. The aggregation of collection data serves to meet the need of libraries in cataloguing and resource sharing, and also provides great opportunities for repurposing the data in Trove. Libraries Australia provides a mechanism for aggregating information about the distributed works of authors and other creators spread across Australian collections. It also plays a role in ‘syndicating’ the aggregated data into discovery services which contribute to the broader information infrastructure.

The value that is provided to member libraries in return for the subscription fee can be summed up as follows:

- Member libraries are able to use all of the ANBD data to meet their copy cataloguing needs, and find all of that data in one place.
- This data includes Table of Contents data from Blackwells and the Library of Congress, which enhances the member libraries’ catalogues.
- Member libraries are eligible to receive and use the Libraries Australia Cataloguing Client.
- Member libraries have unlimited access to WorldCat – and other external databases – with their subscription, giving them a high copy cataloguing hit rate.
- Member libraries will receive assistance in making transitions such as migration to new cataloguing standards (including Resource Description & Access).
- Member libraries are able to make full use of the national interlibrary loan system, LADD (Libraries Australia Document Delivery); if they have a local ILL system they receive support to fully test the interconnection of this system with LADD and therefore with other ILL systems, as part of their subscription.
- Member libraries are able to access support through the Libraries Australia Help Desk, whenever they need it.
- Member libraries are able to offer their end users alerting services and enhanced ILL requesting services through the personalisation features of Libraries Australia Search.
- The catalogue data of member libraries is synchronised to WorldCat in real time, and is therefore visible to a global community.
- The syndication of data to WorldCat means that member libraries are able to take advantage of services such as the OCLC Collection Analysis service, which can assist with collection development and de-selection.
- The catalogue data of member libraries is also exposed globally to other services such as Google.


Underlying system
The system which supports Libraries Australia was redeveloped during 2004 and 2005. That redevelopment lowered the costs of providing the service, and reduced the risks involved in any future system redevelopment by segmenting the system into four modules, namely:

- Libraries Australia Search, which enables member libraries to find and get the resources recorded in the ANBD. This module is currently supported by the TeraText search software from SAIC.
Libraries Australia Cataloguing, a data creation and maintenance service, which enables member libraries to create and edit relevant data, and also enables managers of the ANBD to obtain, convert and maintain relevant data. This module is currently supported by the CBS software from OCLC in Leiden.

Libraries Australia Document Delivery, an interlending and payment service, which enables member libraries to manage the process of requesting, shipping and paying for loans and copies. This module is currently supported by the VDX software from OCLC in Sheffield.

Libraries Australia Administration, which enables member libraries to manage their user account information and register for the Record Export Service (RES) and the Products Service. This module is currently based on the Open LDAP directory software.

The metadata which is aggregated by Libraries Australia is being repurposed in the Library’s flagship national discovery service (Trove). In this way, Libraries Australia enables Australians to access resources held in Australian library collections and Australian publications available online.

Between 2006 and 2009 there were two versions of the Libraries Australia Search service: a “subscription” version and a “free” version. The free version was decommissioned at the end of 2009 because its function was subsumed by Trove. For the moment, the subscription version includes “personalisation” features for end users, such as alerts and enhanced requesting. Member libraries have previously been encouraged to make these features available to their end users, but the Division envisages that in the future they will be provided by Trove. Consequently, further development of the Libraries Australia Search service will focus on the needs of member libraries, including any specific search needs of cataloguing and interlibrary loan staff. However, in the longer term Libraries Australia Search and Trove will share a common back-end based on a common Lucene database.

**Strategies for July 2010 to June 2012**

**Strategy 1.1: Encourage and strengthen affiliation and usage of Libraries Australia**

**Background**

The Australian library community benefits from a high level of participation in Libraries Australia. It is in the interests of all libraries for services such as shared cataloguing and interlibrary loan to have the widest possible reach.

Widespread participation will also assist the National Library’s ability to support the service by maintaining revenue at or above the current level.

Libraries Australia has implemented its new subscription model, although some member libraries are still transitioning to their new subscription levels. Member libraries are now charged in a manner which is equitable (in the sense that it is based on intrinsic library size) and which does not discourage high levels of use by the libraries and their users. The new subscription model was implemented for TAFE libraries in July 2010. The subscriptions for university libraries were updated in July 2010 to reflect their updated library budgets. It is planned to review the application of the model for each sector at three year intervals. Thus, the subscriptions for state, territory and public libraries were reviewed in early 2011 for implementation in July 2011, and those for special libraries will be reviewed in late 2011 for implementation in July 2012. It is incumbent on the Division to communicate and explain these changes to all affected libraries.
Specific activities

In support of this strategy, the Division will:

- extend and maintain partnerships with member libraries, promoting the Libraries Australia value proposition;
- encourage more special libraries and school libraries to join Libraries Australia;
- maintain the subscription model to ensure an equitable sharing of financial support for the service;
- ensure that the availability and performance of the four system modules are maintained at target service levels;
- provide support to member libraries that are implementing their own ISO-compliant ILL systems, by testing the interoperability of these systems with LADD;
- review the options for the continuing operation and development of LADD including the strategic direction of the VDX software and the future of the ISO ILL protocol;
- support the Libraries Australia Advisory Committee, and seek the Committee’s involvement and advice;
- support a project to to improve both awareness and accessibility of hidden collections in special libraries;
- communicate well with member libraries through:
  - an improved Libraries Australia web presence, including an online member forum;
  - professional and friendly Help Desk service;
  - provision of written information through the annual report, other LAAC papers, and the e-newsletter;
  - support for sectoral events;
  - regular visits to state user groups and sectoral library groups; and
  - an attractive programme each year for the Libraries Australia Forum;
- continuously improve the documentation for Libraries Australia services including online user manuals and tutorials;
- educate member libraries on how to maximise use of and derive the greatest benefit from Libraries Australia; and
- support the training agents.

Strategy 1.2: Enhance Libraries Australia services

Background

To meet the changing needs of the member libraries, the Libraries Australia software will be regularly improved and extended. These changes will eliminate sub-optimal features and develop new service offerings. The changes will be made to all four modules of the software.

It has been assumed that, for the duration of this Plan, Libraries Australia Cataloguing will continue to be supported by the CBS software, and that Libraries Australia Document Delivery will continue to be supported by the VDX software.

During 2008 and 2009 the Libraries Australia Search service performed below its target levels for availability and performance. These performance shortcomings were due to some inherent limitations in the software application. A number of configuration changes, including the removal of the relevance ranking of search results, have now improved search availability and performance, but some limitations remain in meeting target update volumes. The medium to
long term strategy to address this issue will be to migrate this Libraries Australia module to the same software platform (Lucene and Solr) that is used by Trove.

The Division is leading the National Library’s collaboration with state and territory libraries in the Reimagining Libraries Open Borders Project. One aim of this Project is to identify options for the state and territory libraries to leverage off Trove in developing their own discovery services. The Division plans to work with some of these libraries to explore the feasibility of enabling these local services, including the local catalogues, to become “views” of Trove. Achievement of this outcome may depend on improvements to the software which would allow data of local significance to be recorded and managed within or in association with the ANBD.

Following these investigations, the Division would need to undertake a thorough analysis prior to any decision to implement a “local catalogue” service offering. It is unlikely that such an offering would be made before July 2013.

Specific activities

In support of this strategy the Division will, during the period of this Plan:

- scope, specify and commence the project to migrate Libraries Australia Search to the Solr/Lucene platform;
- deliver one more release of Libraries Australia Search on the existing platform (TeraText), including changes required by RDA, and priority improvements to the Record Export Service and Products (examples: further develop authority record products; allow Division staff to have more control over the running of products);
- continue to add “deep links” to local catalogues from Libraries Australia Search;
- through our contract with OCLC Leiden, deliver two new releases of Libraries Australia Cataloguing, including the completion during 2011 of the CBS “Job Manager”, which will enable staff of the Libraries Australia office to undertake tasks which currently must be performed by IT staff at the National Library or at OCLC Leiden;
- through our contract with OCLC Leiden, undertake configuration changes to CBS to support RDA;
- make priority changes to the Record Import Service (examples: complete the development of the Global Refresh Fileset; develop capability to produce RIS files from OAI and Harvest Control List gathering of local data);
- through our contract with OCLC, deliver two new releases of Libraries Australia Document Delivery;
- integrate the Australian Libraries Gateway and the Inter Library Resource Sharing Directory, and explore the feasibility of automated migration of the data into the WorldCat Registry;
- modify the Australian Libraries Gateway to support authenticated access to e-resources in Trove;
- deliver enhancements to Libraries Australia Administration and explore with selected member libraries (particularly state and territory libraries) the feasibility of enabling their local discovery services, including their local catalogues, to become “views” of Trove.
**Strategy 1.3: Improve the ANBD**

The Division’s knowledge of bibliographic standards, and its commitment to quality assurance, are valued by the member libraries and ensure a high quality Libraries Australia service. Consequently, the Division will continue to develop the ANBD in terms of its coverage, timeliness and consistent data quality.

Data can be contributed to the ANBD by three methods:

1. The Libraries Australia Cataloguing Client;
2. The Record Import Service; and
3. The simple Web Cataloguing Form.

The Division has no preference between these methods, but encourages each member library to identify the combination of methods that best suits its local needs. The Division does aim, however, for:

- widest possible coverage of member library data, including data about e-resource subscriptions and original materials;
- timeliness in data contribution; and
- contribution of high priority authority data from state and appropriate university libraries.

Following the implementation of CBS 5.0 in mid 2010, the Division is able to support a new contribution method, based on the SRU Record Update protocol. Where local library systems cannot support this protocol, the Division will aim to test the implementation of the OCLC Synchronisation Gateway with interested member libraries. This Gateway widget will support near real-time update of the ANBD, with the data exported from the local system in conformity to the SRU Record Update protocol.

During 2008 the Division undertook a major survey of the practices and capabilities of member libraries in contributing to the ANBD. As expected, many member libraries are not contributing all of their eligible data. Where data is being contributed, it is sometimes not contributed in a timely manner. It will benefit all member libraries if the scope and timeliness of data contribution can be improved.

It is likely that, in the first half of 2012, the national libraries of several countries, including Australia, will implement the new cataloguing rules called Resource Description and Access (RDA). The Division will encourage the member libraries of Libraries Australia to adopt this new standard.

The Collections Management Division of the National Library is coordinating collaboration with state and territory libraries in the Reimagining Libraries Description and Cataloguing Project. One aim of this Project is to ensure that collection summaries, finding aids and lists are contributed to the ANBD and/or Trove. Usually this will involve the contribution of a brief collection-level record to the ANBD, with the full text of the collection summary indexed in Trove. The Resource Sharing and Innovation Division will work closely with this Project.

The strong public use of Trove has led to responses from a number of end users, who have discovered that the holdings data in Trove (and therefore in the ANBD) is not up to date. A key
issue here is the fact that the outcomes of public library de-selection activities are often not reported to the ANBD. During the life of this strategic plan, the Division will give priority to identifying and implementing efficient solutions to this problem.

It should also be noted that the Division undertakes constant maintenance of the ANBD so that it can be used with confidence by the member libraries.

Specific activities

In support of this strategy, the Division will, during the period of this Plan:

- support the implementation of the new cataloguing standard RDA (Resource Description and Access) so that member libraries can create and exchange RDA-compatible data (the timing of this action will depend on decisions by national libraries concerning the RDA implementation date);
- in following up the ANBD Survey, work systematically with member libraries to discuss gaps in their data contributions, and to implement measures to eliminate or reduce these gaps (priority to be given to state and territory libraries, then university libraries);
- improve reporting of holdings deletions from public library de-selection activities;
- work with member libraries, OCLC, and if necessary system vendors, to encourage the implementation of the SRU Record Update protocol, or the WorldCat Synchronisation widget, so that local systems of member libraries can update the ANBD as a real-time background task;
- continue to refine the database matching and merging algorithms;
- continue to configure and refine the operation of the offline duplicate record removal facility;
- continue to configure and encourage use of the service which supports batch mode import of authority records (including an examination of options for loading LC authority files);
- improve ANBD authority files: for example, link bibliographic and authority records, and merge name and name/subject headings, with consequent improvements to bibliographic records; and
- continue to enrich the ANBD by harvesting additional value-added content, such as Table of Contents data from the Library of Congress and other sources.

Strategy 1.4: Maintain and develop business relationships

The quality of services that are offered to member libraries depends in part on the business relationships that the Division develops. For example, these relationships can extend the range of cataloguing data and the range of document supply options that are available to the member libraries.

Of particular importance is the relationship with OCLC, which involves:

- software support by OCLC for Libraries Australia Cataloguing (CBS) and Libraries Australia Document Delivery (VDX);
- a network licence allowing member libraries to make unlimited use of WorldCat as a supplementary source of cataloguing data;
- contribution of the ANBD to WorldCat, with data synchronisation; and
- participation of member libraries in OCLC governance.
The Division will continue the relationship with New Zealand’s library network (Te Puna), especially in the support of trans-Tasman interlibrary lending. It is also planning a strengthened relationship with the British Library Document Supply Service, so that LADD users can easily forward requests to BL DSC.

Specific activities

In support of this strategy, the Division will:
- maintain regular contact with the OCLC offices in Ohio, Leiden and Melbourne, with information sharing leading to joint activities where needed;
- implement system interoperability and contractual arrangements to enable LADD users to forward requests to BL DSC;
- investigate and where possible implement inter-operability improvements;
- support Australian representation on the Asia Pacific Regional Council of OCLC, and participate in forums organised by the Council;
- maintain an active interest in the RLG Programs and Research activities of OCLC;
- contribute to, and exploit, the Virtual International Authority File;
- maintain the relationship with Te Puna;
- maintain and develop relationships with key document supply agencies such as Infotrieve; and
- maintain dialogue with library system vendors.

5. E-RESOURCE CONSORTIA

Service overview

Electronic journals and databases now represent a major part of the value which libraries offer to their users. Australian libraries subscribe to a variety of such electronic resources, sometimes through sectoral or state-based consortia. However, many Australians do not have access to the e-resources they would like to use, because they are not affiliated with a library which subscribes to them. Increased availability of e-resources will contribute to the social, educational and economic well-being of the community.

The Division manages and co-ordinates two consortia for licensing of e-resources: the NSLA E-Resources Consortium, and Electronic Resources Australia (ERA).

The short to medium term goals for these consortia are:
- to assist the member libraries to obtain the best possible prices for e-resource licences; and
- to assist the member libraries to achieve aligned and favourable licence conditions.

The National Library also has two long term aspirations with respect to the e-resources that are licensed by the Australian library community:
- to give every Australian resident the right to access a core set of e-resources, free to the user but paid for by the Australian library community, assisted where possible by direct Government funding; and
- to allow Australian library users to easily discover and link through to those e-resources which they are entitled to access by virtue of their multiple library memberships.
Achieving the second of these aspirations will be a major undertaking for the Discovery Services, discussed below.

Achievement of the first aspiration will necessarily be incomplete unless the National Library, and other library stakeholders, can lobby successfully for Government funding of a core set – either covering the whole Australian community, or covering a specific sector such as schools. The Library has already discussed this matter with some government agencies, including the Department of Broadband, Communications and the Digital Economy and Department of Education, Employment and Workplace Relations.

NSLA E-Resources Consortium

The Division manages the NSLA E-Resources Consortium on behalf of National & State Libraries Australasia (NSLA). The members of the Consortium are Australian state and territory libraries and the National Library of Australia. Members may also elect to act on behalf of one or more public libraries in their state or territory.

The Division negotiates and signs licence agreements on behalf of members, makes payments to the vendors and recovers these from the Consortium members. The licences are renewed, where appropriate, towards the end of each calendar year.

The NSLA E-Resources Consortium updated its Business Plan for the period November 2010 to October 2012. A Committee, representing all members of the Consortium, meets quarterly to make decisions and recommendations about operational issues. In addition, Reimagining Libraries Project 7 (“Collaborative Collections”) is providing advice on the longer term strategies for the Consortium.

In the first half of 2010 the NSLA E-Resources Consortium developed selection criteria for a “core set” of e-resources suitable for take-up by all NSLA member libraries. Using these criteria, it then identified a proposed core set. The Division then sought from vendors “whole of sector” prices for each product in the core set. Based on the information provided, NSLA Chief Executive Officers decided on a phased approach to implementing whole-of-sector licences with nine products to be implemented in the first phase.

Electronic Resources Australia (ERA)

Following the 2003 recommendations of the Senate Inquiry into the Role of Libraries in the Online Environment, the National Library convened a series of forums with representatives from library groups and various consortia. These forums led to the establishment of Electronic Resources Australia (ERA). The Library began to manage ERA in early 2006, calling the first ERA open tender in August 2006, with the results announced in May 2007.

Because of the prior existence of other consortia, in the short to medium term ERA is likely to be of most benefit to small libraries (school, special and some public libraries). ERA will work to improve coverage for a basic set of e-resources, with the long term goal of achieving a situation in which every Australian library user can access the basic set through any library with which they are affiliated.
ERA has a Business Plan covering the period July 2009 to June 2011. By December 2010 over 1,100 libraries from all sectors were taking subscriptions through ERA, either individually or through consortia. At that time the ERA Product Panel included 25 products, from eight participating vendors. In December 2010, ProQuest chose not to renew its Head Deed leaving only one product remaining in the Australian News and Business category. To fill this gap, vendors on the multi-use list will be invited to propose appropriate new products.

In May 2010, ERA made a submission to the House of Representatives Inquiry into School Libraries and Teacher Librarians in Australia Schools. Among other things, the submission advocated Government funding of $4.8M per year for a “core set” of e-resources relevant to school libraries.

Currently, ERA comprises a set of arrangements for particular products which may approximate the “core set” which we wish to make available to every Australian. The arrangements involve individual libraries electing to "opt-in" by referencing the relevant Head Deed in their contracts signed directly with the vendors.

The Division has investigated the option of the National Library negotiating with vendors as a single consortial entity, resulting in a single contract with each vendor. This is the approach adopted by New Zealand’s national consortium Electronic Purchasing in Collaboration (EPIC). While there may in theory be administrative savings for vendors and participating libraries in such a model, discussions to date have not provided a compelling case for taking this approach. This approach would require a greater staffing resource in the Division, requiring an administration fee to be added to the process for users. Vendors have downplayed the possibility of administrative savings and do not favour an approach which creates more distance between them and the subscribing libraries. Consequently the Division will not pursue this model during the life of this Strategic Plan.

The discussions which occurred in early 2010 have however identified a number of improvements that can be made in the pricing process, including the abolition of the “product poll” which to date has been complex and confusing for participating libraries.

The operation of ERA involves:

- maintaining a “multi-use list” panel of e-resource vendors, including annual advertisements on AusTender to provide an opportunity for new potential suppliers to apply for inclusion, or for existing members to modify their product packages;
- reviewing the ERA product offerings annually in consultation with the ERA Executive Committee and undertaking select tenders for new products;
- managing the procurement and establishment of Head Deed agreements with these vendors, which include licence conditions and annual subscription prices;
- promoting ERA and the product offerings to Australian libraries and supporting the promotion of e-resources to their users;
- evaluating the performance of the consortium; and
- operating under a governance framework.

**Licence conditions**

Ideally, libraries that licence e-resources wish to do so under conditions such as the following:
Resource Sharing & Innovation

Strategic Plan

Last updated 13/04/2011
Page 18 of 24

Strategies for July 2010 to June 2012

**Strategy 2.1: Increase participation in the NSLA and ERA products**

Both consortia will benefit from a high level of take-up of the licensed products. Greater participation will bring the consortia closer to achieving whole-of-sector and national prices, which will lead to cost savings for all participating libraries. “Whole-of-sector” prices have been offered for some products by some ERA vendors. Acceptance of such offers would mean that all Australian libraries in a given sector (e.g., school libraries, NSLA libraries, university libraries) would be able to offer the product in question to their users. However, initial testing has revealed that these prices have been set at levels which do not encourage any whole-of-sector take-up.

**Specific activities**

In support of this strategy, the Division will:

- update the Business Plans for ERA and the NSLA E-Resources Consortium;
- streamline the ERA pricing process, including the abolition of “product polls”;  
- promote the ERA products widely to appropriate libraries;
- facilitate training by vendors in use of the products; and
- review the ERA product offerings annually.

**Strategy 2.2: Establish a framework that supports national licensing**

While the Division will not pursue the “single contract” model during the life of this Strategic Plan, it will undertake a number of activities designed to strengthen the national licensing framework. The Division will:

- in consultation with other stakeholders such as ALIA, continue to lobby relevant government agencies for funding to support a core set, either for the whole community of for a specific sector such as schools;
if necessary, liaise with state authorities in an effort to obtain coverage for all libraries in a particular sector in that state;
- negotiate with vendors in an effort to obtain more attractive whole-of-sector prices;
- in particular, obtain whole-of-NSLA prices for the new NSLA core set, with the objective of taking up those licences wherever possible;
- aim to bring the NSLA E-Resources Consortium and ERA closer together in a national approach to providing access to e-resources; and
- participate in Project 7 (“Collaborative Collections”) of the NSLA Reimagining Libraries initiative.

6. DISCOVERY SERVICES FOR THE PUBLIC

Overview of existing services

The Library began to provide free online discovery services for the public within two years of establishing its web site. These services provide a single point of discovery for a range of information resources that are held in multiple Australian collections.

These services are now undergoing a process of integration and transition. In November 2009 the Library released Trove, a new service which will over time replace all of the separate discovery services that were developed between 1997 and 2008. In December 2009, two of these services (Libraries Australia Free Search Service, and Register of Australian Archives and Manuscripts) were decommissioned. The stand-alone Australian Newspapers service was decommissioned in November 2010 making Trove the sole interface to Australian newspapers.

The stand-alone services and programs supported by the Resource Sharing and Innovation Division in late 2010 were:

- Picture Australia, which provides a single point of discovery for pictures that have been digitised by Australian collecting institutions (libraries, museums, archives and other institutions). By December 2010 it was providing access to over 1.8 million images
- Music Australia, which presents a specialised subset of the ANBD (over 260,000 records in December 2010) containing a wide range of resources relating to Australian music, including books, pictures, biographical information, music scores, sound recordings and websites. The underlying database is a copy of this subset, presented with its own online interface
- Australian Research Online (formerly known as the ARROW Discovery Service), which in December 2010 was providing access to more than 550,000 Australian research outputs, including theses; preprints; postprints; journal articles and book chapters. The service aggregates the metadata from participating Australian university repositories, and selected government research repositories.

The Division also supports a program that is aggregating metadata describing people and organisations - the People Australia program. This program provides unique identifiers for these parties. It supports the presentation, within Trove, of a virtual web page for each person or organisation encompassed by the program (over 880,000 entries in December 2010). Users of Trove can navigate from these web pages to entries in the ANBD for works by and about the person concerned, and can also navigate to biographical information, and organisational profile information, contained in other web-based services.
Trove is not only a discovery service: it is also an annotation and content contribution space. Any of the information resources exposed by Trove can be the subject of user activity to add comments and tags, create lists, correct the text of digitised newspaper articles, and contribute content including digital photographs.

**Strategies for 2009-2011**

**Strategy 3.1: Expand and enhance Trove**

The cornerstone strategy for the next two years is to expand and enhance Trove, and to integrate the stand-alone services which Trove is replacing.

Trove has already improved the discovery experience for the Australian public by providing access from one place to the resources previously discoverable via eight discovery services. It is already providing access to a significantly greater range of resources from a wider range of sources, including more full-text content. It has enhanced ease of discovery through improved relevance ranking, refinement by facets, and grouping of related items. It is now allowing users to engage with the content by supporting annotation (tagging and commenting) services and the creation of lists.

The Resource Sharing and Innovation Division is the business owner for Trove, with responsibility for ensuring that data is successfully harvested for the service. However, other Divisions and Branches of the Library take responsibility for particular content as appropriate. This responsibility includes defining the policies for data collection, and maintaining some of the stakeholder relationships.

The Resource Sharing and Innovation Division and the other relevant Divisions and Branches of the Library will collaborate to:

- actively seek out additional content, whether structured metadata or full text, and with an emphasis on data held by Australian partner institutions; and
- identify appropriate external targets for inclusion in the new service.

This Project will assist the State and Territory libraries to improve their discovery services as part of the “Open Borders” Project of the NSLA Reimagining Libraries initiative.

**Specific activities**

During 2011, the following activities are planned to expand and enhance Trove:

- Adding a range of interface and search improvements in response to the outcomes of a usability review;
- Integration of Australian Digital Theses (ADT) into Trove;
- Scoping and developing a standards-based programmer API, in consultation with identified stakeholders (this development will allow other discovery services to use Trove as a target);
- Adding a base set of journal article content, including indexing data from the APAIS (Australian Public Affairs Information Service) and AMI (Australian Medical Index)
services, and e-resource metadata from Informit, with support for access to the full text of e-resources in cases where that can be easily implemented;

- Extending access to e-resources to encompass additional vendors (e.g. Cengage Gale) and additional access options (e.g. authentication of users); and

- Completing harvesting and delivery improvements required to integrate Picture Australia and Australian Research Online.

Subject to corporate approval in light of other National Library priorities, further development of Trove during the life of this Plan may involve activities such as:

- Conducting a pilot of interlibrary loan requesting by end users for selected libraries;
- Investigating and if possible implementing a mobile device interface.

**Strategy 3.2: Encourage a high level of use of Trove**

The Division will encourage a high level of use and user acceptance of the service.

In early 2010 the Division, in collaboration with the National Library’s Communications and Marketing Branch, developed a Marketing Plan for Trove. In April 2010 a publicity campaign saw significant media exposure and a temporary surge in the use of Trove.

In June 2010 an online User Forum was introduced, based on the vBulletin forum software. This Forum primarily aims to support dialogue between users.

**Specific activities**

In support of this strategy, the Division will:

- take opportunities to raise awareness of Trove on the part of potential user groups such as education communities and creative industries;
- encourage users to register, annotate the content and participate in the online forum;
- undertake social networking activities (eg blogging and twittering) to promote Trove;
- speak about Trove at conferences and seminars;
- use the Trove information pages to encourage more user-generated content; and
- acknowledge individuals for their contributions.

**Strategy 3.3: Develop the communities of contributors and partners**

The Division will aim to attract new Trove contributors, and new partners whose content can be accessed as external targets.

**Specific activities**

In support of this strategy, the Division will:

- engage with a wide range of collecting institutions, including archives, museums and galleries, universities and other research repositories;
- engage with other institutions and projects that are active in the aggregation of collection metadata, such as the ANDS-funded “Cultural and Historical Collections Metadata Exchange” Project, which involves 15 museum partners;
- give priority to the addition of data from major national cultural and collecting institutions;
establish an online contributor space so that data contributors can communicate more easily with each other and with the Division’s staff;
convene an annual face-to-face Trove contributors meeting or workshop;
identify suitable contributors through knowledge built up from other National Library activities (example: Community Heritage Grants);
extend the scope of the Trove Harvester software to support harvesting by SRU and OpenSearch/RSS; and
investigate making contribution more flexible to overcome any barriers to contributing data.

7. THE WIDER COLLABORATIVE AGENDA

The Resource Sharing and Innovation Division is engaged in a broad set of collaborative activities which support the Library’s Directions and also support the future development of the three service bundles discussed above.

Reimagining Libraries
The Division will play a strong role in leading or supporting several of the projects that form part of the Reimagining Libraries initiative of National & State Libraries Australasia. These projects will strengthen the other services offered by the Division. Among these projects are:

- Open Borders (led by the Division) which will plan how NSLA member libraries will make use of Trove in their own discovery services, and will develop a shared strategy to improve access through Trove to e-resources of the member libraries and Australia’s public libraries;
- Collaborative Collections (led by the State Libraries of Qld and NSW) which will support the efforts to move toward national licensing of e-resources, and including e-resource holdings data, with associated licence information, onto the ANBD;
- Description and Cataloguing (led by National Library of Australia, Collections Management Division) which identify more streamlined and efficient processes for creating metadata and for sharing this metadata through Libraries Australia and Trove;
- Delivery (led by the State Library of WA) which will provide an impetus towards transforming document delivery and interlibrary loan, and will pick up the aims of the ‘Rethinking Resource Sharing Reference Group’ which has investigated the possibility of end user requesting and home delivery of items; and
- Virtual Reference (led by the State Library of Victoria) which will plan for the next generation of virtual reference services and define connections with other services such as the Trove User Forum.

Party Infrastructure Project
Between January 2010 and June 2011 the Division is undertaking a project in partnership with the Australian National Data Service (ANDS). The Australian Research Data Commons (ARDC) Party Infrastructure Project, funded by ANDS, aims to enable improved discovery of research outputs and data through persistently identifying Australian researchers and research organisations.
This project will extend the software established for the People Australia program, to build a sustainable architecture to identify researchers and research organisations well into the future. In particular, the Project will:

- extend the People Australia Data Administration module to enable distributed name authority control by information officers in research organisations;
- improve the automatic matching algorithms currently in use;
- provide standard machine interfaces to data about researchers and research organisations;
- extend the data standards used in the service to cater for the record standards used in the research sector;
- provide access to persistent identifiers for researchers and research organisations; and
- enhance the People Australia Data Administration module to allow data managers in universities to manage the data that they contribute about researchers.

**Standards activities**

The Division will continue to be involved in standards development and implementation activities, where these relate to the Division’s services. These activities will include:

- participation in Standards Australia Committee IT/19;
- participation in the development of the Encoded Archival Context (EAC) standard, which supports the exchange of biographical information;
- participation in the Australian Committee on Cataloguing; and
- encouraging the implementation of protocols such as SRU Record Update, which will improve efficiencies and timeliness in the exchange of bibliographic data.
Evaluation of the Libraries Australia Forum

Introduction

The fifth annual Libraries Australia Forum was held at the National Library of Australia on 20 October 2010. The Forum was part of three days of related events in Canberra, including the Libraries Australia Advisory Committee (LAAC) meeting, a Libraries Australia Forum Welcome Reception and four library tours on 19 October and the Australian Committee on Cataloguing (ACOC) Seminar on 21 October.

There were 227 attendees for the Forum this year, compared to 146 in 2009. The Forum was promoted via the regular channels – messages to the Libraries Australia and Libraries Australia Document Delivery mailing lists, at State User Group Meetings, the Libraries Australia Home Page and on the Libraries Australia space on Ning.

Keynote speakers at the Forum were Constance Malpas, Program Officer, Research Libraries Group, OCLC, who spoke about the “future of collecting”, with a focus on how academic libraries could share digitised collections and Katie Birch, Portfolio Director, Delivery Services, OCLC, who spoke about upcoming trends in document delivery and inter-library loan.

Among other highlights, the Forum heard presentations on special libraries and the sector from both Australian Library Journal editor & LAAC Special Libraries representative Ann Ritchie; and the National Library’s Acting Director of Collaborative Services, Laurel Paton. Acting Assistant Director-General, Resource Sharing and Innovation Debbie Campbell reported on the year’s activities in Libraries Australia, with a special focus on the new Trove search service. NSLA Consortium & National Licensing manager Nikki Darby gave an update on Electronic Resources Australia and Trove manager Rose Holley presented the results of the Research Library Group’s Social Metadata Working Group.

Attendees were asked to fill out an evaluation form prior to departing from the Forum. An opportunity to fill the form in online after the event was also offered, and responses were accepted for the following two weeks. This report summarises these evaluation forms. Statistical responses can be found at Attachment A of this report.

In total, 97 responses were collected from attendees. This equates to a 43% return rate from the total number of attendees, slightly up from 40% in 2009.
Overall Meeting

Approximately 90% of respondents to the evaluation rated the Forum overall as above average or excellent. Around 9% rated it as average, and 2% as below average. This is similar to previous years, with a slightly higher polarisation: with more rating it is as excellent, but also a few below average ratings.

Comments included:

‘NLA always puts together a good conference and this was no exception. Well thought through and structured’

‘Overall a good meeting, really enjoyed tour, too’

‘Good reminder of the variety of projects LA is involved in and reminder that NLA is there to help all kinds of small libraries with all kinds of problems’

‘A number of issues to take back to work – why aren’t our holdings on Libraries Australia? Why aren’t we part of ERA? We all twitter to promote our databases and build a community of learning, what can we learn from Rose [Holley’s] research?’
Parallel sessions

The parallel sessions were well received, with 73% of respondents rating the sessions as either excellent or above average. Twenty-five per cent rated these sessions as average, while only 2% rated sessions as below average and none considered the presentations to be of poor quality. These were the highest approval ratings for parallel sessions since the Forum’s inception.

There were three interactive sessions that ran concurrently. One on Interlibrary Loans covering current ILL issues, overview of LADD developments; another on Cataloguing and CBS Futures which discussed current and future plans for Libraries Australia cataloguing services, as well as how catalogued records appear in Trove. A third workshop, presented by the Libraries Australia Help Desk, discussed Enhanced Requesting, and took questions from the audience. The cataloguing session was the most well-attended, and the comments reflected the usefulness of this session to technical services professionals.

Some comments about the sessions were;

‘Cataloguing and CBS [was the session I liked the most] as this is my area and I like to see what’s coming’

‘ILL interactive session [was the best] because there was an opportunity for troubleshooting questions to be answered’

‘Enhanced requesting and meeting LA Helpdesk [offered a] great opportunity for communication’

Quality of Presentations

Eighty-eight percent of respondents rated the quality of presentations either above average or excellent, while 1% of respondents considered the presentations to be

Figure 2: Parallel sessions – 2008-2010
below average. The most common suggestion for improvement was to reduce the amount of acronyms and “jargon” used by presenters.

The evaluation form was slightly updated this year to ask which presentations attendees found most useful. While every session was considered among the most useful by at least one respondent, the four most popular sessions were Constance Malpas’ presentation on the future of collecting (45% of respondents); Rose Holley’s presentation on social metadata (36%); the parallel session on cataloguing (33%) and the presentation of the future of Document Delivery by Katie Birch (29%).

Some comments:

‘OCLC presentation [was a] high calibre and professional presentation’

‘Constance Malpas’ [presentation] was very professional and the international perspective was very useful’

‘Social media presentation was fun, interesting and informative’

![Figure 3: Quality of presentations – 2008-2010](image_url)

**Meeting Content**

Eighty-five percent of respondents rated the meeting as above average or excellent, while 15% rated the meeting content as average, and 3% rated the meeting content as below average.

Comments received:

‘I have learnt a lot. I also realised that trends and directions of our own library fit well into the context and activities of the wider community of libraries and info/resources’

‘The Forum provided a well balanced approach; cataloguing/metadata of access to resources; trends and new developments’
‘Good, varied program’

Figure 4: Meeting Content – 2008-2010

Opportunities for discussion

Seventy-eight percent of respondents rated the opportunities for discussion as above average or excellent, while 25% rated these opportunities as average, and 4% below average. Comments indicated that while most found the parallel sessions well structured for discussion, a few attendees felt this could be improved.

Some example comments include:

‘The parallel session worked really well. There was lots of discussion. I look forward to notes from the session’

‘Be good to explore more opportunities for questions and answers. Maybe these could/should be run in smaller sessions where people are more comfortable speaking up.’

Figure 5: Opportunities for discussion – 2008 - 2010
Meeting Venue/Catering

In contrast to the fifty percent rating in 2009, 85% of respondents found the venue excellent or above average, and 14% found it average. Some Forum attendees commented on the air-conditioning being ineffective early in the morning, however, most comments were overwhelmingly positive. Comments regarding the catering included:

‘The catering was excellent’

‘Food and refreshments were good and the weather was above average’

Analysis of Attendees

The Forum is held in a different metropolitan city each year according to the following arrangement. On rotation, a choice is made between Adelaide, Brisbane, Hobart and Perth. In the alternate year, a choice is made between Sydney, Canberra and Melbourne. The location does influence attendance, as shown by the comparison chart for each state and territory:

The Forum was held in Sydney in 2008, Hobart in 2009, and Canberra in 2010.
The attendance by library sector is shown in this comparison chart for 2008 to 2010:

![Attendees by sector](image)

*Figure 8: Analysis of attendees by Library Sector – 2010*

**Acknowledgement**

This paper was prepared by Alison Dellit.

**Recommendation**

The Libraries Australia Advisory Committee to note the report.

**Contact:**
Debbie Campbell  
Director Collaborative Services  
dcampbel@nla.gov.au

17 January 2011
Attachment A:

Overall analysis of 2010

<table>
<thead>
<tr>
<th></th>
<th>Excellent</th>
<th>Above Average</th>
<th>Average</th>
<th>Below Average</th>
<th>Poor</th>
<th>Total responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Meeting</td>
<td>33</td>
<td>54</td>
<td>9</td>
<td>2</td>
<td>0</td>
<td>97</td>
</tr>
<tr>
<td>Parallel sessions</td>
<td>22</td>
<td>49</td>
<td>24</td>
<td>2</td>
<td>0</td>
<td>97</td>
</tr>
<tr>
<td>Quality of presentations</td>
<td>31</td>
<td>53</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td>Meeting content</td>
<td>26</td>
<td>55</td>
<td>14</td>
<td>3</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td>Opportunities for discussion</td>
<td>27</td>
<td>14</td>
<td>16</td>
<td>2</td>
<td>0</td>
<td>96</td>
</tr>
<tr>
<td>Meeting venue/catering</td>
<td>37</td>
<td>45</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>97</td>
</tr>
</tbody>
</table>
### 2008-2010 comparison

<table>
<thead>
<tr>
<th>Overall meeting</th>
<th>Excellent</th>
<th>Above Average</th>
<th>Average</th>
<th>Below Average</th>
<th>Poor</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses 2008</td>
<td>31</td>
<td>97</td>
<td>57</td>
<td>1</td>
<td>0</td>
<td>186</td>
</tr>
<tr>
<td>Percentage 2008</td>
<td>16.70%</td>
<td>52.20%</td>
<td>30.50%</td>
<td>0.50%</td>
<td>0.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Responses 2009</td>
<td>15</td>
<td>37</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>57</td>
</tr>
<tr>
<td>Percentage 2009</td>
<td>26.30%</td>
<td>64.90%</td>
<td>8.80%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Responses 2010</td>
<td>33</td>
<td>54</td>
<td>9</td>
<td>2</td>
<td>0</td>
<td>97</td>
</tr>
<tr>
<td>Percentage 2010</td>
<td>34.02%</td>
<td>55.67%</td>
<td>9.27%</td>
<td>2.06%</td>
<td>0.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parallel sessions</th>
<th>Excellent</th>
<th>Above Average</th>
<th>Average</th>
<th>Below Average</th>
<th>Poor</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses 2008</td>
<td>19</td>
<td>77</td>
<td>74</td>
<td>16</td>
<td>2</td>
<td>185</td>
</tr>
<tr>
<td>Percentage 2008</td>
<td>10.30%</td>
<td>41.60%</td>
<td>40.00%</td>
<td>8.60%</td>
<td>1.10%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Responses 2009</td>
<td>9</td>
<td>27</td>
<td>18</td>
<td>1</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>Percentage 2009</td>
<td>16.40%</td>
<td>49.10%</td>
<td>32.70%</td>
<td>1.80%</td>
<td>0%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Responses 2010</td>
<td>22</td>
<td>49</td>
<td>24</td>
<td>2</td>
<td>0</td>
<td>97</td>
</tr>
<tr>
<td>Percentage 2010</td>
<td>22.70%</td>
<td>50.52%</td>
<td>24.70%</td>
<td>2.06%</td>
<td>0.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>
### Quality of Presentations

<table>
<thead>
<tr>
<th></th>
<th>Excellent</th>
<th>Above Average</th>
<th>Average</th>
<th>Below Average</th>
<th>Poor</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Responses 2008</strong></td>
<td>18</td>
<td>100</td>
<td>51</td>
<td>3</td>
<td>1</td>
<td>185</td>
</tr>
<tr>
<td><strong>Percentage 2008</strong></td>
<td>18.40%</td>
<td>54.10%</td>
<td>27.60%</td>
<td>1.60%</td>
<td>0.05%</td>
<td>100.00%</td>
</tr>
<tr>
<td><strong>Responses 2009</strong></td>
<td>21</td>
<td>30</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>57</td>
</tr>
<tr>
<td><strong>Percentage 2009</strong></td>
<td>36.80%</td>
<td>52.70%</td>
<td>10.50%</td>
<td>0%</td>
<td>0%</td>
<td>100.00%</td>
</tr>
<tr>
<td><strong>Responses 2010</strong></td>
<td>31</td>
<td>53</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td><strong>Percentage 2010</strong></td>
<td>32.60%</td>
<td>55.80%</td>
<td>14.70%</td>
<td>1.05%</td>
<td>0%</td>
<td>100%</td>
</tr>
</tbody>
</table>

### Meeting Content

<table>
<thead>
<tr>
<th></th>
<th>Excellent</th>
<th>Above Average</th>
<th>Average</th>
<th>Below Average</th>
<th>Poor</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Responses 2008</strong></td>
<td>29</td>
<td>87</td>
<td>64</td>
<td>3</td>
<td>0</td>
<td>183</td>
</tr>
<tr>
<td><strong>Percentage 2008</strong></td>
<td>15.80%</td>
<td>47.50%</td>
<td>35.00%</td>
<td>1.60%</td>
<td>0.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td><strong>Responses 2009</strong></td>
<td>15</td>
<td>32</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>56</td>
</tr>
<tr>
<td><strong>Percentage 2009</strong></td>
<td>26.80%</td>
<td>57.10%</td>
<td>16.10%</td>
<td>0%</td>
<td>0%</td>
<td>100.00%</td>
</tr>
<tr>
<td><strong>Responses 2010</strong></td>
<td>26</td>
<td>55</td>
<td>14</td>
<td>3</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td><strong>Percentage 2010</strong></td>
<td>27.40%</td>
<td>57.90%</td>
<td>14.70%</td>
<td>3.16%</td>
<td>0%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>
### Opportunities for discussion

<table>
<thead>
<tr>
<th></th>
<th>Excellent</th>
<th>Above Average</th>
<th>Average</th>
<th>Below Average</th>
<th>Poor</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses 2008</td>
<td>28</td>
<td>81</td>
<td>64</td>
<td>9</td>
<td>2</td>
<td>183</td>
</tr>
<tr>
<td>Percentage 2008</td>
<td>15.30%</td>
<td>44.30%</td>
<td>35.00%</td>
<td>4.90%</td>
<td>1.10%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Responses 2009</td>
<td>14</td>
<td>26</td>
<td>14</td>
<td>2</td>
<td>0</td>
<td>56</td>
</tr>
<tr>
<td>Percentage 2009</td>
<td>25.00%</td>
<td>46.40%</td>
<td>25.00%</td>
<td>3.60%</td>
<td>0%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Responses 2010</td>
<td>27</td>
<td>14</td>
<td>16</td>
<td>2</td>
<td>0</td>
<td>96</td>
</tr>
<tr>
<td>Percentage 2010</td>
<td>28.13%</td>
<td>14.60%</td>
<td>16.70%</td>
<td>2.10%</td>
<td>0%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

### Meeting venue/Catering

<table>
<thead>
<tr>
<th></th>
<th>Excellent</th>
<th>Above Average</th>
<th>Average</th>
<th>Below Average</th>
<th>Poor</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses 2008</td>
<td>51</td>
<td>73</td>
<td>51</td>
<td>10</td>
<td>1</td>
<td>186</td>
</tr>
<tr>
<td>Percentage 2008</td>
<td>27.40%</td>
<td>39.20%</td>
<td>27.40%</td>
<td>5.40%</td>
<td>0.05%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Responses 2009</td>
<td>12</td>
<td>17</td>
<td>24</td>
<td>4</td>
<td>0</td>
<td>57</td>
</tr>
<tr>
<td>Percentage 2009</td>
<td>21.00%</td>
<td>30.00%</td>
<td>42.00%</td>
<td>7.00%</td>
<td>0%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Responses 2010</td>
<td>37</td>
<td>45</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>97</td>
</tr>
<tr>
<td>Percentage 2010</td>
<td>38.10%</td>
<td>46.40%</td>
<td>14.40%</td>
<td>0%</td>
<td>0%</td>
<td>99%</td>
</tr>
</tbody>
</table>
Cloud-sourcing Research Collections: Managing Print in the Mass-digitized Library Environment

Constance Malpas

Program Officer
OCLC Research
Cloud-sourcing Research Collections: Managing Print in the Mass-digitized Library Environment
Constance Malpas, for OCLC Research

© 2011 OCLC Online Computer Library Center, Inc.
Reuse of this document is permitted as long as it is consistent with the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 (USA) license (CC-BY-NC-SA):
http://creativecommons.org/licenses/by-nc-sa/3.0/.
January 2011

OCLC Research
Dublin, Ohio 43017 USA
www.oclc.org

OCLC (WorldCat): 695086590

Please direct correspondence to:
Constance Malpas
Program Officer
constance_malpas@oclc.org

Suggested citation:
Contents

Acknowledgments ........................................................................................................... 7

Executive Summary ...................................................................................................... 8

Introduction ................................................................................................................... 13
  Premise ...................................................................................................................... 14
  Methodology .............................................................................................................. 14
  Scope of Analysis ...................................................................................................... 15

Summary of Findings .................................................................................................... 17
  Shared Digital Repository Profile: HathiTrust ............................................................. 17
  Shared Print Repository Profile: ReCAP ................................................................. 32
  Model Consumer Profile: NYU ............................................................................. 45

Shared Print Provision: Assessing the Options ............................................................... 50
  Expanding the Scope of Shared Service ................................................................. 50
  Assessing Market Maturity ..................................................................................... 51
  Alternative Service Providers .................................................................................. 52
  Optimizing Existing Infrastructure ......................................................................... 55

What is It Worth? Putting a Price on Shared Collection Services .................................... 58

Conclusions and Recommendations ............................................................................. 64

Appendix I. HathiTrust Cost Rationale ....................................................................... 67
Appendix II. Cloud Library Service Agreements: ReCAP as Shared Print Repository ...... 71
Figures

Figure 1. Growth of HathiTrust Digital Library collection (June 2009 - June 2010) ............ 18
Figure 2. Projected growth of HathiTrust Digital Library (June 2010 - June 2020) ............ 19
Figure 3. Primary document types of titles in HathiTrust Digital Library (June 2010) ........ 20
Figure 4. Distribution of HathiTrust Digital Library titles by document type (June 2009 - June 2010) ........................................................................................................ 21
Figure 5. Subject distribution of titles in HathiTrust Digital Library (June 2010) ............... 22
Figure 6. Distribution of titles in HathiTrust Digital Library by subject and copyright status (June 2010) ........................................................................................................ 27
Figure 7. Top ten categories of public domain content in HathiTrust Digital Library (June 2010) ........................................................................................................ 29
Figure 8. System-wide distribution of library holdings for titles in HathiTrust Digital Library (June 2010) ........................................................................................................ 31
Figure 9. Distribution of ReCAP holdings by contributor (July 2010) .............................. 33
Figure 10. Growth in titles duplicated in ReCAP and HathiTrust Digital Library (September 2009 - June 2010) ......................................................................................... 34
Figure 11. Primary document types of titles duplicated in ReCAP and HathiTrust Digital Library (June 2010) ........................................................................................................ 37
Figure 12. Subject distribution of Hathi titles held in ReCAP (June 2010) ....................... 38
Figure 13. Comparative scope of shared digital and shared print repository collections (June 2010) ........................................................................................................ 40
Figure 14. Titles duplicated in ReCAP and the HathiTrust Digital Library (June 2010) ....... 42
Figure 15. System-wide distribution of library holdings for Hathi titles in ReCAP (June 2010) 44
Figure 16. Growth in coverage of NYU Bobst holdings in HathiTrust Digital Library (June 2009 - June 2010) ........................................................................................................................................ 46

Figure 17. NYU Bobst titles duplicated in ReCAP and HathiTrust Digital Library (September 2009 - June 2010) ............................................................................................................. 47

Figure 18. NYU Bobst titles duplicated in UC SRLF and HathiTrust Digital Library (June 2009 - June 2010) .................................................................................................................. 53

Figure 19. Comparison of potential shared print provision options for NYU Bobst Library (June 2010) ........................................................................................................................................ 54

Figure 20. NYU Bobst titles duplicated in ReCAP partner libraries and HathiTrust Digital Library (June 2009 - June 2010) .......................................................................................................... 56

Figure 21. Percentage duplication of titles held in ARL libraries and HathiTrust Digital Library (June 2009 and June 2010) ........................................................................................................ 62
Into being
The clouds condense, when in this upper space
Of the high heaven have gathered suddenly,
As round they flew, unnumbered particles—
World’s rougher ones, which can, though interlinked
With scanty couplings, yet be fastened firm,
The one on other caught.

Lucretius *De rerum natura*, Book V
trans. William Ellery Leonard (1921)

Acknowledgments

The Cloud Library project emerged out of a series of discussions that began with Carol Mandel, Jim Neal, John Wilkin and Jim Michalko in 2009. These individuals provided leadership and vision that guided all the work that followed.

Library staff from New York University, Columbia University, the New York Public Library and Princeton University participated in a variety of meetings, conference calls and e-mail exchanges that helped to give shape to the project. The Andrew W. Mellon Foundation contributed financial support under a grant ably administered by Chuck Henry at the Council on Library and Information Resources (CLIR).

Michael Stoller, Bob Wolven, Zack Lane, Matthew Sheehy, Marvin Bielawski and Eileen Henthorne made essential contributions to the project, not least in helping to compile ReCAP holdings data for inclusion in our analysis. Kat Hagedorn and Jeremy York provided expert technical and operational support from Hathi. Jenny Toves ensured that WorldCat data extractions were available on schedule.

I am grateful to Jim Michalko, John Wilkin and Paul Courant for their many thoughtful questions and suggestions about the data analysis and interpretation. Lorcan Dempsey and Brian Lavoie also provided insights and helpful methodological guidance along the way.

Particular thanks are due to Roy Tennant and Bruce Washburn, who provided expert programming support over the course of this project and routinely produced small miracles, and to Patrick Confer for his diligent editorial work in preparing the final report.
Executive Summary

The Cloud Library project was jointly designed and executed by OCLC Research, the HathiTrust, New York University’s Elmer Holmes Bobst Library, and the Research Collections Access & Preservation (ReCAP) consortium, with support from The Andrew W. Mellon Foundation. The objective of the project was to examine the feasibility of outsourcing management of low-use print books held in academic libraries to shared service providers, including large-scale print and digital repositories.

The following overarching hypothesis provided a framework for our investigation:

- The emergence of a mass-digitized book corpus has the potential to transform the academic library enterprise, enabling an optimization of legacy print collections that will substantially increase the efficiency of library operations and facilitate a redirection of library resources in support of a renovated library service portfolio.

From this, a number of research questions emerged:

- What is the scope of the mass-digitized book corpus in the HathiTrust Digital Library and to what degree does it replicate print collections held in academic research libraries?

- Can public domain content in the HathiTrust Digital Library provide a suitable surrogate for low-use print collections in academic libraries?

- Is there sufficient duplication between shared print storage repositories and the HathiTrust Digital Library to permit a significant number of academic libraries to optimize and reduce total spending on local print management operations?

- What operational gains might be obtained through a selective externalization of collection management activities?

Based on a year-long study of data from the HathiTrust, ReCAP, and WorldCat, we concluded that our central hypothesis was successfully confirmed: there is sufficient material in the
mass-digitized library collection managed by the HathiTrust to duplicate a sizeable (and
growing) portion of virtually any academic library in the United States, and there is adequate
duplication between the shared digital repository and large-scale print storage facilities to
enable a great number of academic libraries to reconsider their local print management
operations. Significantly, we also found that the combination of a relatively small number of
potential shared print providers, including the Library of Congress, was sufficient to achieve
more than 70% coverage of the digitized book collection, suggesting that shared service may
not require a very large network of providers.

Analysis of the distribution of subject matter and library holdings represented in the
HathiTrust Digital Library and shared print repositories further confirmed that the digital
corpus is largely representative of the collective academic library collection, suggesting a
broad potential market for service. A further positive finding was that monographic titles in
the humanities constitute the greatest part of the mass-digitized resource, which may
indicate that some relatively under-resourced disciplines will begin to benefit from a digital
transformation that has already powered enormous innovation in the sciences. As detailed
below, we also found that substantial library space savings and cost avoidance could be
achieved if academic institutions outsourced management of redundant low-use inventory to
shared service providers.

Our findings also revealed some important obstacles and limitations to implementing changed
print management practices in the current library operating environment. The following are
among the most important constraints we identified:

- The proportion of public domain content in the HathiTrust Digital Library is relatively
  small (approximately 16% of titles in June 2010) and typically represents material that
  is not widely held in the library system; as a result, the number of libraries that might
  hope to reduce local print management costs for these titles through negotiated
  agreements with the HathiTrust and shared print providers is quite low. Moreover, the
  age and subject distribution of titles in the public domain is not representative of
  academic research collections as a whole. In sum, the public domain corpus as
currently defined by U.S. copyright law cannot be considered a viable surrogate for
any academic print collection.

- While significant duplication was found between the HathiTrust Digital Library and
  multiple large-scale library storage collections, it was apparent that no single print
  storage repository could offer coverage sufficient to enable significant space savings or
  cost avoidance for a given client library. Put another way, effective shared print
  storage solutions will depend upon a network of providers who will need to optimize
  holdings as a collective resource.
The absence of a robust discovery and delivery service based on collective print storage holdings is an impediment to changed print management strategies, especially for digitized titles in copyright.

It is our strong conviction, based on the above findings, that academic libraries in the United States (and elsewhere) should mobilize the resources and leadership necessary to implement a bridge strategy that will maximize the return on years of investment in library print collections while acknowledging the rapid shift toward online provisioning and consumption of information. Even, and perhaps especially, in advance of any legal outcome on the Google Book Search settlement, academic libraries have a unique opportunity to reconfigure print supply chains to ensure continued library relevance in the print supply chain. In the absence of a licensing option, online access to most of the digitized retrospective literature will be severely constrained. Demand for print versions of digitized books will continue to exist and libraries will be motivated to meet it, but they will need to do so in more cost-effective ways. In the absence of fully available online editions, full-text indexing of digitized in-copyright material provides a means of moderating and tuning demand for print versions and should facilitate the transfer of an increasing part of the print inventory to high-density warehouses. Viewed in this light, shared print storage repositories could enable a significant and positive shift in library resources toward a more distinctive and institutionally relevant service portfolio.

Our study assessed the opportunity for library space saving and cost avoidance through the systematic and intentional outsourcing of local management operations for digitized books to shared service providers and progressive downsizing of local print collections in favor of negotiated access to the digitized corpus and regionally consolidated print inventory. As detailed in the report that follows, the organizational change required to achieve these gains is likely to be substantial and challenging to implement. Yet, the opportunity costs of inaction may prove even greater than the risks of enacting shared print management regimes. Many of the positive transformations that academic library directors hope to achieve in the next decade or so will require a fundamental shift in collections management. The scope and scale of change that is possible may be judged by these key findings:

- As of June 2010, the median rate of duplication between titles held by university libraries in the U.S. Association of Research Libraries (ARL) and the HathiTrust Digital Library exceeds 30%; that is to say, nearly a third of the content purchased by research-intensive libraries in the United States has already been digitized and is preserved in a shared digital repository.

- If the current growth trajectory of the HathiTrust Digital Library is sustained, we can project that more than 60% of the retrospective print collections held in ARL libraries
will be duplicated in the shared digital repository by June 2014. This growth rate far exceeds average annual acquisitions in ARL libraries, suggesting that the digital replication of legacy collections will outpace growth of new physical collections, enabling a transformation in traditional library operations, staffing and space requirements.

- The median space savings that could be achieved at an ARL library if a robust shared print offer were in place today amounts to approximately 36,000 linear feet or the equivalent of more than 45,000 assignable square feet (ASF). These are conservative estimates based on the assumption that holding libraries own a single copy of each duplicated title. Actual space savings could be much greater. In practical terms, this means each library could recover space sufficient for a learning or research commons, media lab, or office space for faculty and visiting scholars.

- The total annual cost avoidance that could be achieved if shared print service provision for mass-digitized books were available today would amount to a figure between $500,000 and $2 million per ARL library, depending on the physical environment (e.g., open stacks on campus or high-density off-site storage) in which the titles would be managed locally.

Academic library directors can have a positive and profound impact on the future of academic print collections by adopting and implementing a deliberate strategy to build and sustain regional print service centers that can meet aggregate demand with aggregate supply. Beyond the obvious operational efficiencies of consolidating low-use, digitized print volumes into shared service collections there is an important strategic advantage to reconfiguring collective inventory that is increasingly devalued as an institutional asset. A proactive effort to rationalize collections that are undergoing a radical phase change from print to digital will enable libraries to achieve a careful and measured wind-down of operations that no longer deliver distinctive value, while continuing to uphold a vital preservation and access mandate.

The shared infrastructure needed to support a broad-based externalization of legacy print management functions is unlikely emerge without directed action and decision-making by leaders in the academic library community. Individuals and organizations interested in advancing these changes are encouraged to consider the following recommendations:

*Library directors and managers can . . .*

- Advocate in favor of licensed access to the mass-digitized resource as part of a comprehensive strategic plan in which the library can reassert its role as a vital part of the academic enterprise.
• Engage directly with faculty and academic officers to communicate a compelling strategy in which selective externalization of traditional functions is demonstrably improving the institution’s ability to fulfill an academic and research mission.

• Support the HathiTrust’s ongoing efforts to expand public access to the mass-digitized book corpus by affiliating with the organization as a content contributor or sustaining partner.

Prospective shared print providers, including managers of large print storage facilities, can . . .

• Proactively build collections that will deliver maximum operational value to external audiences; leverage the collective library investment in mass digitization and the HathiTrust by accelerating the transfer of mass-digitized titles to print preservation repositories.

• Contribute to the establishment of a common service profile by surfacing model agreements and engaging in community dialog about the operational and business requirements of shared service provision.

Research organizations, including OCLC Research, Ithaka S+R, JISC and other similar entities, can . . .

• Advance our collective understanding of the changing profile of demand for legacy print collections in the mass-digitized environment.

• Help to characterize the optimal redistribution of library resources in different regional and national contexts.

Funding bodies, including IMLS, the Mellon Foundation, NEH and others, can . . .

• Provide funding to support the implementation of shared print management through grants to libraries and other organizations to subsidize the direct costs of title selection and processing until such activities are fully subsumed as ongoing library operations.
Introduction

In spring 2009, a group of ARL directors came together to discuss a common set of challenges and opportunities facing university libraries and identify some shared strategies for responding to them. A number of circumstances were converging that appeared to offer some potential relief from critical space pressures in the library and the increasingly burdensome operations associated with managing a large local inventory of low-use print collections.

The seemingly imminent resolution of the Google Book Search settlement was an important motivating factor: academic libraries were confronting the prospect, at once daunting and liberating, of licensed access to a massive aggregation of digitized books from major U.S. research collections. Would such a collection substantially duplicate local print holdings? If so, what consequences might ensue for traditional academic library operations?

At the same time, the emergence of the HathiTrust, a shared digital repository consolidating much of the library-contributed content from the Google Books database, appeared to resolve many of the concerns the library community had regarding long-term stewardship of the mass-digitized book corpus. In combination with the large aggregations of low-use print collections managed in high-density library storage facilities, Hathi might bridge the gap between a well-documented decline in the use of academic print collections and the anticipated shift toward scholarly reliance on full-text electronic resources.

The fact that critical elements of the shared infrastructure needed to effect a large-scale transition from print to electronic research collections were owned and managed by the library community itself gave library directors confidence that the timing and outcomes of this transition could be managed according to the needs of the academic community and not dictated by the business objectives of commercial providers. Were the combined resources of Hathi and large-scale shared print providers already sufficient to mobilize a change in library operations? What was the scope of service likely to be? How much and what kind of value would it need to deliver? Who—which kinds of libraries and in what number—would benefit? These questions were compelling enough to justify a joint research project in which potential service providers and consumers could explore business requirements, service expectations and feasibility of implementation.
The initiative that emerged from these discussions within ARL came to be known as the “Cloud Library” project, because it posited a future in which library collections and services would be sourced from external providers, reducing local infrastructure and operational expenditures in a manner analogous to the cloud-sourced business and computing solutions that now prevail in the commercial and high-tech sectors. Funded by The Andrew W. Mellon Foundation, the project was staffed by a team of investigators from the HathiTrust, the Research Collections Access and Preservation consortium (ReCAP), New York University Libraries, and OCLC Research. This report provides a high-level summary of findings from this project.

Premise

The research questions that motivated this study reflect a conviction shared by all of the participating institutions: the emergence of a mass-digitized book corpus has the potential to transform the academic library enterprise, enabling an optimization of legacy print collections that will substantially increase the efficiency of library operations and facilitate a redirection of library resources in support of a renovated library service portfolio. We started from the presumption that academic libraries will be motivated to transfer resources (space, personnel, and capital) from local print management operations to shared print and digital repositories in proportion to the tangible benefit that cooperative management confers. We were therefore less interested in examining the theoretical advantages of shared service provision than in characterizing the operational gains (space recovery and cost avoidance) that might be obtained through a selective externalization of collection management activities.

Methodology

Between June 2009 and June 2010, a monthly snapshot of records was harvested by OCLC Research from the publicly available HathiTrust metadata repository. These records were machine-processed to extract OCLC numbers and, where necessary, to extract and map alternative identifiers (LCCN, ISBN or ISSN) to valid OCLC numbers. The resulting batch of OCLC numbers was used to extract bibliographic records and holdings data from the WorldCat database each month. These bibliographic master records were then merged with selected Hathi metadata and (starting in September 2009) a sample of associated ReCAP repository customer codes to produce a single, consolidated dataset for analysis.

A master database was built to support analysis of the compiled data, which was programmatically enhanced to support analysis of key attributes of the aggregate collection, including broad subject areas, total library holdings, institutional source of the digitized text and copyright status. This database was enriched each month with successive snapshots of the
Hathi repository, mapped to WorldCat holdings and ReCAP customer codes as described above. By June 2010, the project database comprised 37 million records, representing a longitudinal view of the growing corpus of library-owned titles that are duplicated in print and digital repositories.

**Scope of Analysis**

In the twelve months covered by this project, the HathiTrust Digital Library doubled in size, increasing from approximately 3 million volumes to more than 6 million volumes. On a per-volume basis, the shared digital repository is now larger than the average ARL library collection; the median reported holdings at university-based ARL libraries in 2008 was approximately 3.5 million volumes. Because our analysis of the HathiTrust collection focuses on unique titles (manifestations or editions), rather than physical items, the number of records we compiled each month was somewhat smaller than the number of records in the Hathi metadata repository. Not every volume in the HathiTrust represents an individual book or journal title, and there is at least some duplication in content ingested from different contributors; as a result, the total number of volumes in the Hathi repository is more than the number of titles covered in our analysis. In June 2009, we identified approximately 2 million unique titles in the HathiTrust Digital Library; by June 2010, that number had grown to more than 3.6 million titles. For purposes of comparison, this represents a collection comparable in scope to research libraries in the top tier of the U.S. ARL rankings, based on holdings set in the WorldCat database. Indeed, at the time of writing, the number of unique titles in the HathiTrust Digital Library exceeds the number of titles cataloged and held by many research libraries.

A key goal of this research project was to assess the scope of coverage in shared print and shared digital repositories, with a view to understanding how the combined resources might enable a local reduction in redundant print inventory. For this reason, it was important to understand how much of the print storage collection in ReCAP is duplicated—or is likely to be duplicated—in the HathiTrust Digital Library. As of this writing, the shared ReCAP facility holds more than 8 million items contributed by the three partner libraries. Since the ReCAP collection is not currently visible as a discrete set of holdings in WorldCat, and building a union catalog of ReCAP holdings was beyond the scope of this project, we based our analysis on a representative sample of ReCAP holdings supplied by Columbia University and NYPL. Taken collectively, Columbia and NYPL’s ReCAP holdings amount to more than 75% of current inventory and this was deemed to be sufficient for our analysis.

The sample supplied to us included a broad range of materials managed under 14 different ReCAP customer codes, each representing a different set of request and circulation rules. The large size and broad scope of the sample gave us reasonable confidence that findings from our
analysis could be generalized across the ReCAP collection as a whole. Storage, selection and transfer protocols at the three partner libraries are based on common parameters (low use monographs; journals duplicated in electronic format), so that the nature, if not the content, of the materials contributed by each is likely to be comparable.

To provide a baseline against which duplication of ReCAP holdings in the HathiTrust Digital Library might be assessed, we periodically compared patterns in the ReCAP sample against other large-scale print storage collections that are more readily subject to analysis in WorldCat. Findings from these analyses are presented below.
Summary of Findings

In this section, the scope and character of holdings in the HathiTrust Digital Library and ReCAP print repository are examined with a view to their potential value in a shared service environment. We first consider the range of holdings in the HathiTrust Digital Library, on the premise that the vast and still expanding scope of the mass-digitized corpus will be a key driver in the transformation of academic library collections and services. We then examine the intersection of titles held in the HathiTrust Digital Library and the ReCAP print repository to assess the degree to which large-scale storage collections might serve as print management hubs, reducing the total cost of preservation and access for low use print resources. Finally, we explore how this shared infrastructure might affect library operations and resource allocations in a research-intensive academic library, using NYU’s Elmer Holmes Bobst library as an exemplar.

Shared Digital Repository Profile: HathiTrust

Over the period of study, the number of volumes in the HathiTrust Digital Library more than doubled, growing from about 3 million items to more than 6.3 million items; the number of titles increased by 90%, from just over 1.9 million titles in June 2009 to about 3.64 million titles in June 2010. Growth was variable from month to month, ranging from a low of about 43,000 new titles in April 2010 to a high of more than 297,000 new titles in November 2009. On average, the number of unique titles in the database increased by about 6% each month. This represents an average increase of nearly 150,000 new titles each month. The ratio of volumes to titles in the repository remained relatively stable at 1.6:1 over the twelve months of this study.
Figure 1. Growth of HathiTrust Digital Library collection (June 2009 - June 2010)

If this rate of growth is sustained, we can expect the HathiTrust Digital Library to rival major research library collections in both size (volumes) and scope (titles) in a matter of a few years. Based on the projections shown below, we can anticipate that the HathiTrust Digital Library collection may be equal in size to Harvard University Libraries (which reported holdings of some 16 million volumes in the 2007-2008 ARL Annual Statistics) by 2013. Within a decade, it could cross the threshold of 30 million volumes, making it larger than the U.S. Library of Congress is today.
Figure 2. Projected growth of HathiTrust Digital Library (June 2010 - June 2020)

For ease of presentation, these projections compare the growth of Hathi to a baseline of constant volume counts at the largest university and non-university ARL collections. Of course, it is reasonable to expect that volume counts for print holdings at these libraries will continue to grow over the next decade; however, the current growth rate of the HathiTrust Digital Library substantially outpaces median annual growth rates at ARL member libraries (approximately 2% of total volume count, based on recent ARL statistics) so we can anticipate that the overlap in digitization of retrospective print holdings will continue to grow faster than the acquisition of new print titles.¹

Understanding the relative distribution of document types in the HathiTrust Digital Library archive is important to characterizing and quantifying its value as a potential surrogate to locally-held academic library print collections. Since the advent of the e-journal transition of the 1990s, university libraries have regarded print versions of dual-format titles as obvious targets for relegation to storage facilities. A major focus of the present study was to determine the degree to which mass digitization of library print collections has resulted in the creation of a digitized book corpus sufficient to enable a similar shift in management of monographic holdings. It is not yet known if the emergence of a large-scale digital book corpus will be sufficient to effect a change in scholarly practice comparable to what has been achieved in the transition from print to electronic journals. Nor is it possible to foresee when, or even if, a legal settlement will be reached that will permit Google to offer universities licensed access to the millions of books that have already been digitized through its
partnerships with academic libraries. While uncertainty about the speed and timing of the format transition for scholarly monographs abounds, we can at least begin to assess the scope and coverage of the academic print collection as it is mirrored in the mass-digitized corpus preserved in the HathiTrust Digital Library.

**Document types**

A vast majority of titles in the Hathi repository represent monographic language-based materials (books). Based on our analysis, books account for 95% of all titles in the HathiTrust Digital Library for which we were able to identify an OCLC number; serial titles comprise approximately 4% of such titles. The remainder of the archive is composed of digitized musical scores, articles, visual resources and the like. While the total volume of non-book and non-journal titles in the archive, as measured in absolute numbers, is impressive (amounting to nearly 50,000 titles in June 2010), these materials collectively represent only about 1% of the Hathi corpus.

![Figure 3. Primary document types of titles in HathiTrust Digital Library (June 2010)](image)

Over the course of our study, an increase in the diversity of document types in the HathiTrust Digital Library has been noted, as indicated by a slight but perceptible shift in proportional distribution of titles. Between June 2009 and June 2010, the relative volume of “other” document types increased from a tenth of a percent (.1%) to a third of a percent (.3%) of all titles in the database. As of June 2010, musical scores account for the vast majority of titles...
in this “other” category. It is not certain what the impact of this trend is likely to be, but one might speculate that a sustained growth in non-book and non-serial titles will be associated with a net decrease in the number of libraries eligible to transfer preservation functions to Hathi, as aggregate library holdings for non-book materials tend to be significantly lower than for book and “book-like” materials. Based on an August 2010 snapshot of the WorldCat database, for example, the average number of library holdings set on an individual monographic title is nine; for musical scores, by contrast, the average number of holdings is four. A shift towards greater representation of non-book and journal content in the archive may meet the needs of current contributors, but it is not likely to support a broader externalization of preservation functions in other libraries.


Because we are primarily concerned with assessing the potential impact of shared digital and print archives on library-managed print collections, and because books continue to represent the single largest cost driver in library operations, the analysis that follows focuses on books and not other library-owned material types.

**Subject distribution**

Individual titles in our dataset were coded with broad and narrow topical descriptors derived from the OCLC Conspectus subject classification. We analyzed the frequency of these codes to determine which subject areas predominate in the digitized Hathi corpus, with the expectation that libraries will adjust print retention policies in view of differing disciplinary...
reliance on physical books. As shown in the chart below, more than 50% of titles in the HathiTrust Digital Library in June 2010 represent content from traditional humanities fields: language and literature, history, philosophy, art and architecture, etc.

Figure 5. Subject distribution of titles in HathiTrust Digital Library (June 2010)

The relative abundance of titles in the humanities (history, language and literature, philosophy) in the HathiTrust Digital Library provides encouraging evidence that mass digitization of library book collections is redressing a long-observed imbalance in the online availability of scholarly resources in the humanities and social sciences, compared to the natural sciences and technology. The HathiTrust’s explicit mandate to increase the educational and research value of mass-digitized books and to improve public access to them should raise library confidence that the vast and still growing aggregation of digitized texts will not only prove satisfactory to students and researchers, but also sufficiently robust to enable a gradual transformation of the library enterprise, as operations shift from locally managed print to collectively managed digital formats.
Books in the humanities typically constitute a significant share of any academic library’s print inventory. While circulation rates for these materials are generally low, they are commonly considered essential to the practice of research and teaching. They have an equally important symbolic value as the embodiment of institutional investment in disciplinary communities that are comparatively “under-resourced” in higher education. Historians are often among the most vociferous critics of any effort to shift physical collections from a central library location to a peripheral shelving or storage annex. Their unease and sometimes outright hostility to well-intentioned strategies for optimizing the distribution of library collections are motivated by deep and praiseworthy concerns about long-term preservation and access to the scholarly record. Until recently, academic libraries have had few options but to retain as much of this low-use but highly valued material on campus as possible; providing direct and unmediated access to print volumes has been the easiest and sometimes the only way to satisfy faculty expectations. The large-scale format transition achieved through mass digitization of these legacy collections has the capacity to transform academic library operations by expanding the range of access options that are available to faculty and students, while simultaneously enabling library managers to make more strategic use of diminishing collections space.

Though smaller in size, other subject-based categories of content represented in the HathiTrust Digital Library are also worthy of note. For example, library owned reference collections (fact books, annual bibliographies, statistical yearbooks, etc.) amount to more than 95,000 titles in the HathiTrust Digital Library. While this constitutes only 3% of the Hathi collection as a whole, it represents a significant potential cost savings for libraries since superseded reference titles are generally regarded as a low print preservation priority; thus, we can imagine that expectations for redundancy in library holdings for these resources might be significantly impacted by replication in the HathiTrust Digital Library. There are more than 20,000 digitized reference titles in the HathiTrust Digital Library that are held in print format in 100 or more libraries. If redundancy in system-wide holdings were reduced to just 15 print copies per title—a figure that recent studies suggest is adequate to ensure survivability of at least one copy for the next one hundred years (Schonfeld, 2009)—a total of more than 20 miles in shelf space might be recovered by libraries.

Government publications are another category of material for which substantial reductions in library print inventory might be achieved, in view of the preservation guarantees provided by the HathiTrust Digital Library. As of June 2010, there are more than 100,000 government documents in the HathiTrust Digital Library collection. More than 40% of these titles are held by in excess of 100 libraries—far more than is required to support the requirements of the U.S. Federal Depository Library Program, for example, and arguably more than is needed to ensure universal access. Because government publications are typically exempt from copyright restrictions, there is every reason to believe that digitized versions will be widely
available, further reducing the need for print inventory. Among titles classified as government documents in the HathiTrust Digital Library, nearly 80% are designated as public domain content. One can easily imagine that many academic libraries will choose to downsize local document collections in favor of online versions; for such institutions, the Hathi preservation services could provide a compelling and cost-effective alternative to local print archiving. Even those libraries that choose to maintain their status as selective depositories could achieve significant cost savings by transferring physical copies of the government publications replicated in the HathiTrust Digital Library to high-density storage facilities.

Additional research is needed to discover what subject areas are included in the “unknown classification” category; given the large number of titles in question (more than 300,000 as of June 2010), this appears to be a fruitful area for study, especially because—as is noted below—more than 20% of titles in this category are in the public domain. Such analysis was beyond the scope of the present study.

Although it was not a focus of our analysis, we did note the presence of many large FRBR work sets in the HathiTrust Digital Library, which suggests some intriguing possibilities not only for discovery services but also for cooperative management and preservation. Thus a library holding a print version of a low-use, in-copyright title might be more likely move it to a cost-efficient high-density facility if it had negotiated with Hathi to provide a link to a public domain digitized surrogate. Another library might opt to withdraw holdings based on levels of duplication in the HathiTrust Digital Library for the associated work set. Our investigation suggests that 5% or more of titles in the Hathi collection (as of June 2010) can be associated with larger work sets. Popular titles like Defoe’s Robinson Crusoe or Swift’s Gulliver’s Travels, as well as classics like Lucretius’ De rerum natura or Homer’s Iliad, are each represented by hundreds of digitized editions in the HathiTrust Digital Library; the long-term preservation of the intellectual work embodied in these manifestations is, to coin a phrase, virtually guaranteed.

It is worth considering that as the number and scope of variant editions in Hathi grows, its value to the academic library community may increase exponentially, enabling the Trust to offer valuable preservation services even to libraries that have contributed no content to the collection. This could significantly increase the market for Hathi preservation and access services and would entail measuring duplication in holdings not on a volume or title level, but on a FRBR work level. In this scenario, Hathi would provide a bridge to facilitate the transition of scholarly practice from print to electronic resources, incrementally reducing demand for, and expectations of, physical proximity to print holdings. Thus, some number of the more than two thousand libraries that hold print editions of Sinclair Lewis’ Babbitt might reasonably opt to shift the locally-held print version to a high-density storage warehouse.
while providing patrons with full-text reading access to a digitized public domain version. Libraries availing themselves of this service would still be “on the hook” for preservation of editions not replicated in the Hathi collection, but could manage those resources more efficiently. In this sense, every library that holds an edition of a work represented in the Hathi repository is in a position to derive some tangible benefit from participation in the network. This has important implications for the future growth of the HathiTrust Digital Library, since the capacity to benefit from participation will increase as the scope of the collection increases to include more widely-held titles and work sets.

**Rights status**

One of the hypotheses that this study set out to test is that the HathiTrust Digital Library represents a potentially rich source of digital surrogates that might, over time, effectively replace a substantial proportion of low-use print collections in academic libraries. It was therefore important not only to examine the size and growth of this corpus over time, but also to consider the degree to which it replicates print holdings in the wider academic library system.

For most of the twelve-month period covered by this study, the relative proportion of in-copyright and public domain content in the HathiTrust Digital Library remained stable, with about 17% of volumes designated as public domain material. This figure increased to about 20% near the end of the project, due in part to a programmatic change in the HathiTrust rights determination algorithm that affected a large number of items ingested earlier in the year. On a per-title basis, a similar distribution was noted over the course of the study, with about 12% of titles designated as public domain content, rising to approximately 16% by the project’s close. As of June 2010, approximately 590,000 titles were designated as “full view” content available for onscreen reading in the HathiTrust platform. About 96% of these public domain titles are books, similar to the distribution pattern noted above for the HathiTrust Digital Library as a whole.

In other respects, the public domain corpus presents significant differences. First and most obviously, titles in the public domain are typically older publications, either published before the 1923 threshold (for U.S. publications) or in the period between 1923 and 1976, when some previously in-copyright titles may be “reborn” as public domain content, either by direct negotiation with the rights holder or by determining that a title eligible for copyright renewal has not been renewed. For this reason, titles in the public domain do not typically represent current scholarship. Some notable exceptions exist, especially where Hathi has negotiated with scholarly publishers to provide public domain access to recent titles and, to a lesser degree, where individual authors have voluntarily released their claim to copyright on titles in the Hathi archive. Nevertheless, the age distribution for the public domain content in Hathi is...
unequivocally skewed toward older titles. *Approximately 80% of the “full view” books in the HathiTrust Digital Library were published prior to 1923; less than 1% were published in the last decade.* By contrast, if we look at the Hathi corpus as a whole, less than 20% of titles were published before 1923; more than 10% were published since 2000. Clearly, the public domain content represents a relatively mature—not to say more authoritative, or more frequently cited—subset of the scholarly record. It is by no means a representative microcosm.

Similarly, if we consider the distribution of public domain content by topical subject area, it is evident that the scope of coverage differs from that of the HathiTrust Digital Library as a whole. For instance, government information constitutes a very small part of the Hathi collection (about 3% of titles in June 2010) but accounts for a disproportionately large share (15%) of titles in the public domain. By contrast, topical areas that are well represented in the mass-digitized corpus, and which typically constitute the greatest part of the academic print collection, account for only a very small part of the public domain resource. Titles in language and literature amount to 25% of the HathiTrust Digital Library as a whole, but represent less than 20% of the public domain corpus. Even more remarkable disparities are evident in Art History and Political Science, disciplines where the monograph is a primary vehicle of scholarly communication. Simply put, the “universal library” of digitized public domain content does not represent a microcosm of the academic print collection.
These findings should not be taken to mean that cooperative agreements aimed at increasing reliance on centralized repositories of digitized public domain content are not worth pursuing. On the contrary, we feel that *there is substantial opportunity for cost efficient reorganization of academic print collections based on the increased availability of public domain content in the HathiTrust Digital Library*. The sheer magnitude of the HathiTrust Digital Library means that even disciplinary resources that comprise a small proportion of the collection as a whole are, in absolute terms, considerable. For example, Philosophy represents a small fraction of the library (6% of all titles in June 2010), but includes a disproportionate number of titles in the public domain: a total of 39,000, or 19% of all titles in this subject area. Language and literature titles are significantly less likely to be in the public domain (13% in June 2010), but the staggering number of titles in this category means that the net yield—some 116,000 titles—is substantial.
For North American libraries especially, the expanding public domain corpus in the HathiTrust Digital Library represents a shared resource of potentially great value. Although it is unlikely to enable a significant change in local print management operations, it unquestionably improves access to a large body of materials that are otherwise relatively difficult to find or obtain. Because out-of-copyright titles are more likely to represent older and more specialized publications, they are most often held in print by only a small number of academic research libraries with a long collecting history (Lavoie and Dempsey, 2010). As a result, these titles are less visible in the library environment and also more difficult to obtain; their relative scarcity means that they are less likely to be available for inter-lending.

The chart below provides a view of the largest subject-based categories of public domain content in the HathiTrust Digital Library, based on title counts in June 2010. These areas appear to represent the greatest near-term opportunity for redirection of library preservation resources, since at least some libraries can be expected to withdraw and replace locally-held physical copies with freely available digital surrogates. At academic and research institutions where off-site and high-density shelving facilities are available, a more systematic and streamlined transfer of low-use print titles from the stacks to storage may be achieved as full-text access eases faculty and librarian concerns about the loss of on-site browsing. Again, the predominance of titles in the humanities is significant, as faculty in history, philosophy and other humanities disciplines are typically the most concerned about relegation of local print inventory. The greater access enabled by full-text provision, in combination with the improved preservation conditions in most off-site facilities, should go some way toward allaying faculty anxiety; if positioned within a larger library strategy for long-term preservation of the scholarly record, it might even embolden faculty to appeal for an accelerated and more aggressive transfer of library holdings off-site.
Figure 7. Top ten categories of public domain content in HathiTrust Digital Library (June 2010)

It’s reasonable to ask if the current distribution of public domain and in-copyright materials in the HathiTrust Digital Library is likely to change over time, as a secondary effect of an increase in the base of content contributors, in response to a programmatic effort to ramp up public domain contributions or even as a result of the ongoing efforts to renegotiate copyright status. The method we used to harvest and process metadata from the Hathi repository makes it difficult to establish any direct correlation between source of contribution and the relative dearth (or abundance) of public domain content. However, as the proportion of public domain content in academic print collections is relatively low—mirroring patterns in historical print production and library collecting behaviors—even a comprehensive effort to digitize and pool these resources is unlikely to result in a significantly different distribution of public domain and in-copyright titles in the HathiTrust Digital Library. One can reasonably expect that the proportion of “full view” titles and volumes in the shared repository will remain stable at about 16% of titles (20% of volumes) for as long as North American research libraries are the primary source of content contribution.

**Distribution of system-wide print holdings**

The distribution of print holdings for titles in the Hathi repository provides some insights into the potential market for digital preservation and access services. We can predict that libraries will be motivated to redirect management operations (and resources) for print holdings that are replicated in the mass-digitized corpus in proportion to their relative
abundance in the system-wide collection, as well as their rights status and online availability. Simply put, the market value of a digital preservation and access offer that enables many libraries to relegate or withdraw a significant volume of redundant inventory will be greater than the value of a similar offer for titles that are of interest to a smaller number of libraries.

An intriguing and potentially significant finding of our analysis is that many titles in the HathiTrust Digital Library are held by relatively few libraries, based on current WorldCat holdings data. **Almost 50% of the 3.64 million titles in the repository as of June 2010 are held by fewer than 25 libraries; 14% are held by fewer than 5 libraries.** Put another way, the market for surrogate preservation services for these titles is limited to a small number of libraries who currently own them and who are (in the near term) unlikely to withdraw them, since they represent distinctive institutional assets. The Hathi preservation service offer for these titles would appear to have less (or more accurately, a different kind of) business value, for the specialized audience of research institutions who collectively “care about” the library long tail. A cooperative service agreement shaped around the shared business needs of the ARL community as a whole, rather than the libraries that hold these titles, would possibly provide a means of broadening the base of service and reducing the cost burden for individual Hathi partners. If these relatively rare materials were explicitly marketed as a common-pool resource, cooperatively managed by members of the ARL community, the number of stakeholders prepared to commit resources to Hathi might be enlarged.
At the farthest end of the library long tail are titles held by a single institution, for which a redistribution of preservation investment seems most challenging. In June 2010, the HathiTrust Digital Library included more than 190,000 such titles, representing about 5% of the collection as a whole. These resources are similar in format and content to uniquely-held print materials examined in previous studies, with an abundance of grey literature, pamphlets, non-English (especially East Asian) titles and, above all, a great number of dissertations and theses (Connaway, O’Neill and Prabha, 2007). Most of the titles in this latter category were contributed by the University of Wisconsin. The rights distribution for Hathi titles with a single holding library is not much different from other titles; approximately 10% are in the public domain. These resources may have great scholarly value, but there is no evidence that they are more accessible as a result of digitization.

The abundance of titles in the HathiTrust Digital Library that are relatively scarcely held should not obscure the fact that there is opportunity for significant library space recovery associated with de-duplication of low-use titles for which aggregate library supply exceeds projected demand. As of June 2010, there are at least 25,000 titles archived in digital format by Hathi for which collective library print holdings per title exceed 1,000 libraries; more than 900 titles in the HathiTrust Digital Library are held in print by more than 2,500 libraries. It is difficult to imagine a preservation scenario that would require this level of redundancy in the system-wide print collection. There is considerable debate and discussion in the library community regarding optimal thresholds of duplication in print.
collections. One widely-cited study posits that a minimum of 15 unsecured copies of any given title are needed to ensure survivability of a single copy after one hundred years, assuming typical library loss rates (Schonfeld, 2009). This model presumes an as yet non-existent network of print preservation guarantees expressed by individual libraries. However, if even a relatively small number of copies are secured in preservation-quality print repositories, a carefully planned strategy to reduce system-wide print inventory is not only theoretically possible but operationally feasible.

As the quality and conditions of use for mass-digitized books continue to improve, as they surely will for titles in the shared Hathi repository, one can imagine that shared print repositories will emerge as an acceptable and even preferred alternative to local management of the mass-digitized book corpus.

**Shared Print Repository Profile: ReCAP**

A key hypothesis that this study was designed to test is that there is sufficient duplication between shared print storage repositories and the HathiTrust Digital Library to permit a significant number of academic libraries to optimize and reduce total spending on local print management operations. There are at least four library print storage facilities in the United States with holdings in excess of 5 million volume-equivalents that might be supposed to rival the HathiTrust Digital Library in scope of coverage (Payne, 2007). If adequate duplication between these individual repositories and the HathiTrust Digital Library already exists (or can be attained), one can imagine a scenario in which client libraries would contract with a regional print repository and with Hathi for preservation and access services, progressively externalizing some portion of local print management operations. For the purposes of this study, we focused in particular on the Research Collections Access and Preservation consortium (ReCAP) facility, which manages low-use collections deposited by Columbia University, the New York Public Library (NYPL) and Princeton University. In June 2010, the ReCAP collection included more than 8.5 million items.
Using sample data provided by Columbia University and NYPL, we examined rates of duplication in ReCAP holdings compared to the HathiTrust Digital Library. Deposits from Columbia and NYPL account for more than 75% of items accessioned by ReCAP, which was considered sufficient for analysis. We were supplied with a sample of approximately four million item-level records (about two million from each library), which were then processed to extract OCLC numbers for matching against the project database. Data from Columbia were processed and merged into the project database in September 2009; data from NYPL were added in March 2010. For this reason, it is not possible to provide a representation of longitudinal changes in coverage of ReCAP holdings replicated in the HathiTrust Digital Library. Moreover, since our ReCAP sample data represents a snapshot of the repository holdings at a discrete point in time, any growth in duplication that we are able to report reflects changes in the composition of the Hathi collection and not new accessions in the ReCAP facility. A further limitation is that because no centralized bibliographic database of ReCAP holdings exists, it is not possible to compare the number of ReCAP titles in Hathi to the number of ReCAP titles as a whole.

Despite these challenges, the data we were able to compile and analyze provide some useful insights. *Between September 2009 and June 2010, the number of ReCAP titles in our sample that could be matched to titles in the HathiTrust Digital Library more than doubled, from fewer than 300,000 titles to nearly 700,000 titles.* There are a number of factors contributing to this growth, including some refactoring of code in November which
allowed us to map more of the Columbia data to Hathi records, and the addition of the NYPL data in March. It is clear, however, that the rapid pace of growth in the HathiTrust Digital Library also resulted in a net increase in the number of titles that could be matched.

![Graph showing growth in titles duplicated in ReCAP and HathiTrust Digital Library (September 2009 - June 2010)](image)

**Figure 10. Growth in titles duplicated in ReCAP and HathiTrust Digital Library (September 2009 - June 2010)**

Our analysis suggests that the ReCAP storage collection mirrors a significant portion of the digitized corpus archived in the HathiTrust Digital Library; as of June 2010, nearly a fifth (19%) of titles preserved in digital format by the HathiTrust are also preserved in print format by ReCAP. On the surface of things, this may seem like a surprisingly low figure, given our initial premise that the large digital and print preservation repositories were likely to duplicate one another to a large extent. Indeed, we anticipated that the Hathi and ReCAP collections would overlap to a much greater degree, in part because libraries contributing content to the HathiTrust Digital Library were initially drawing on titles digitized from their own offsite storage collections. It seemed reasonable to believe that the digitized collection of titles from storage collections would have a higher probability of being duplicated in ReCAP (or any other large library storage facility) than in an average academic library’s circulating collection.

It is possible that a more comprehensive analysis of ReCAP holdings, including titles deposited by Princeton University would result in a somewhat higher Hathi duplication rate. Since Princeton deposits amount to a relatively small part (about 20%) of the total ReCAP collection, however, it is unlikely that a more comprehensive analysis would result in a
substantially different figure. A more probable explanation for the lower than anticipated duplication rate between ReCAP and Hathi is that the scope and character of the large storage repositories from which much of the mass-digitized corpus was initially sourced may differ substantially from the holdings on deposit in ReCAP. Farther below, we explore this thesis by comparing the profile of the ReCAP collection against a few other large-scale depositories.

With these caveats in mind, it is worth considering the potential business value of the ReCAP collection as it mirrors the digitized book collection, on the assumption that an increasing number of academic libraries will seek to externalize print management and preservation in coming years. At the time this project commenced, it was generally believed that the digitized Google Books corpus would be made available as a licensed resource, hastening the trend toward externalization of collection management functions in academic libraries. A year later, the likely outcome of the Google Book Search settlement is still unknown, causing us to question whether university libraries will be motivated to outsource preservation of mass-digitized titles in the absence of a comprehensive licensed access option. Yet if the timeline for the digital transition is still uncertain, it is unquestionably the case that academic libraries are being compelled to reconsider the traditional print collection and service portfolio, which was largely dependent on locally managed inventory (Michalko, Malpas and Arcolio, 2010). As a strategic reserve, the ReCAP collection and other similar large-scale depositories could thus offer real value even to non-contributing libraries.

In operational terms, the value of a shared print reserve is potentially far greater than traditional inter-lending and reciprocal borrowing arrangements, if shared service agreements for guaranteed access and preservation are in place. For example, an institution like NYU might find it more cost-effective to purchase guaranteed, just-in-case access to print resources managed in a preservation repository than to retain local copies of low-use titles in a legacy collection. In the context of a formal service agreement, a library’s decision to withdraw local holdings in favor of cooperative preservation and access arrangements would serve a dual purpose of limiting the institution’s exposure to risk while reducing the long-term costs of managing local and even remotely stored inventory.

To understand the degree to which a repository like ReCAP might provide print collection management services scoped around the mass-digitized corpus, it is important to compare not only the relative size of the potential service collection but also its scope and range.

**Document types**

As noted above, the emergence of a mass-digitized book corpus presents enormous opportunity for a positive transformation of library service in the academic sector. Substantial
operational efficiencies have been achieved in library management of the journal literature as a result of format migration and it is not unreasonable to hope that a similar gain can be achieved for legacy monographic collections. Print book collections are a primary cost driver in academic libraries; while journals occupy a disproportionate share of library space on a per-title basis, the operational expenses associated with acquiring, cataloging and serving monographic collections are substantially higher on a per-unit basis. More pertinently, the long-term carrying costs associated with managing monographic collections have remained largely unchanged. While format migration has enabled many university libraries to shift print journal back-files into more cost-effective storage facilities, low-use print book collections still occupy prime campus real estate, at great expense.

If a shared print service collection is to provide maximum value in the mass-digitized book environment, it is obviously important that it include a very large number of monographs that are also represented in shared digital preservation repositories like Hathi. A potential shared print provider like ReCAP would ideally offer print preservation and access services for a significant number of monographic titles in the mass-digitized corpus and deliberately promote and extend this service collection as a source of distinctive value and utility.

The value of a shared monographic collection of this kind would be different and arguably even greater than that offered by a print journal archive, since uncertainties about the long-term demand trajectory for print books (post-digitization) are likely to sustain a broader and more profitable market for service. Profitability in this context is most likely to be measured in terms of increased efficiency in the academic library enterprise; the marginal gain for cooperative management of books will, at least for a time, be greater than for print journals. This is simply a reflection of the fact that libraries have already made significant strides in lowering the costs of managing the journal literature; the incremental gain that might be achieved by further externalizing journal management is less than is possible (and desirable) for books. For this reason, it is encouraging to find that ReCAP already holds a substantial number of mass-digitized books that could form the kernel of a shared service collection.
From a purely pragmatic perspective, implementing shared collection services for a large body of print books may also be somewhat easier than would be the case for serials, where validation of local holdings can be onerous and costly. It is improbable that prospective customers of a shared monographic collection would expect (or pay for) page verification and collation of holdings on a large scale. If required, it could nevertheless be carried out more rapidly and at a lower cost per title for books than for journals.

Subject distribution

Our examination of the Hathi repository found a preponderance of titles in literature, linguistics, history and other humanities disciplines. We consider this a positive finding, since academic library holdings typically include a large share of humanities titles that occupy a correspondingly large share of the library’s physical space. If a significant space savings is to be gained through cooperative management of legacy print collections, it is therefore important that shared service collections include a similarly large share of such titles. Happily, we find that the subject distribution of mass-digitized titles in the ReCAP facility mirrors the distribution of the Hathi corpus as a whole.
This suggests that libraries seeking to “outsource” management of low-use print collections by increasing institutional reliance on shared digital and regional print reserves can realistically expect to transfer preservation and access operations for large monographic collections in the humanities to shared service providers like ReCAP, if appropriate service-level expectations are met. It is worth noting that while the disciplinary scope of such an arrangement will be important in building a market for shared services, the business value of the agreements will ultimately be determined by the actual space savings and cost avoidance that can be obtained. A shared print service offer that enables only a modest impact on local operations will likely fail to mobilize sufficient resources to ensure sustainability.
Beyond the extant scope and scale of the prospective shared print collection, additional factors must be considered in evaluating if it is fit for service at scale. The most important of these is its relative availability to an external clientele. This will be determined by both prevailing access provisions and prospective demand.

**Rights status**

It is important to assess the relative distribution of public domain and in-copyright content in print preservation repositories like ReCAP, since we can anticipate that demand patterns and preservation expectations are likely to be different for titles that are freely available online and those that are subject to more restrictive authentication regimes. For titles in copyright, especially, it is essential that sufficient stock be maintained on a regional or consortium level, as physical copies will remain an important distribution format for some time to come. Based on the findings of this study, we believe that cooperative access and preservation agreements that address the ongoing need for a library print supply chain for in-copyright, digitized books are an essential part of the emerging shared service environment. Indeed we consider the absence of a collective strategy to build a shared print infrastructure that can meet this need will ultimately expose academic libraries to great risk, as the operational focus shifts away from local management of purchased inventory. Finally, from a purely pragmatic point view, it would appear that shared service provision based on an “insurance only” model, where access to print versions of digitized titles is intentionally restricted to exceptional circumstances—for example, when a print version of a digitized public domain title is expressly required—is unlikely to affect the mobilization of library resource needed to sustain shared print repositories.

A simple illustration will suffice to show that a shared print agreement limited to titles in the public domain can deliver only modest benefit to the academic library community. As noted above, a relatively small part of the mass-digitized corpus in Hathi is available as public domain content. Based on our June 2010 snapshot of the Hathi repository, we estimate that this public domain resource amounts to about 600,000 titles, or approximately 16% of the collection as a whole. If ReCAP were to craft a shared print offer around this public domain resource, providing on-demand access to print versions and an assurance of long-term print preservation, it could at best hope to offer a service collection of about 100,000 titles. A small number of academic research libraries might step forward and commit some ongoing financial support for the long-term care of these materials, even without the assurance that this new investment would be offset by a gain in local operational efficiency. But both the size of the pooled resource and the potential audience for such a service are so small that the total impact would only be marginal to the library enterprise as a whole. The economic value of the shared resource in this scenario is further and more fundamentally constrained by the fact that titles in the public domain are less widely
held than in-copyright titles, so that the efficiencies that might be obtained by consolidating physical holdings as a pooled resource are comparatively slight.

By contrast, a ReCAP shared print service offering that included mass-digitized titles currently under copyright could deliver significant value to a large number of academic libraries in the region. In the next section, we consider NYU as an exemplar in the broad market of potential consumers of shared collection services. Here it suffices to observe that as academic libraries look for opportunities to reduce expenditures on operations that can be delivered more effectively at lower cost, it is inevitable that investments in local management of low-use, legacy print collections will come under close and critical scrutiny. *Even, and perhaps especially, in advance of an eventual licensing agreement that will enable libraries to retire local print collections in favor of digital aggregations—a transition that is both deeply feared and fervently desired by scholars and librarians alike—it is imperative that academic administrators begin to plan for a library future in which management operations are selectively and strategically shifted outside the local institution, to larger regional and consortial interests.* Unless this transition is proactively managed by library directors and supported by the academic institutions they serve, there is a strong probability (not to say certainty) that the legacy print collections we have long cultivated as institutional assets will eventually be regarded local liabilities. Unless the collective value of
these resources is accounted for and memorialized in new sets of inter-institutional agreements, responsibility for the preservation of these resources will likely devolve to a handful of research institutions not adequately equipped or empowered to assume a “permanent” stewardship role.

Availability of repository holdings

Materials on deposit in the ReCAP repository are subject to a variety of access rules imposed by the depositing libraries and library units. Thus, books deposited by the Avery Architectural and Fine Arts Library at Columbia University may be subject to circulation and lending restrictions that make them less available than materials deposited by the main Nicholas Murray Butler Library. We used location codes harvested from the ReCAP sample data to examine the relative availability of print versions of titles in the HathiTrust Digital Library, with the expectation that availability of repository collections is certain to be a key factor in negotiating shared print agreements with non-contributing partner libraries. Just as the availability of digitized content in the HathiTrust Digital Library is constrained by intellectual property rights enshrined in copyright law, the availability of print repository holdings is constrained by access rules imposed by owning libraries. Understanding the scope of these constraints is essential to assessing the feasibility of a truly scalable approach to shared service provision.

Fourteen different location codes were included the ReCAP sample data we analyzed; thirteen are associated with Columbia University campus libraries. The chart below reveals the distribution by location or “customer code” of titles in the ReCAP sample that could be matched to digitized titles in the HathiTrust Digital Library in June 2010. Note that because there is some duplication in collections on deposit in the ReCAP collection, the number of ReCAP holdings replicated in Hathi (714,955 volumes) is greater than the number of titles (679,401) that are held in common by both ReCAP and Hathi. As shown below, ReCAP deposits from Butler Library (CU Standard) and NYPL account for the majority of holdings. This is a positive finding, since ReCAP holdings from these libraries are largely unrestricted and therefore potentially in scope for a shared print service agreement. As of June 2010, the ReCAP collection included nearly 600,000 unrestricted titles that mirror content in the HathiTrust Digital Library; this represents a significant pool of resources that might be marketed as a shared print collection. Put another way, almost 90% of the ReCAP collection that is potentially in scope as a surrogate service collection for mass-digitized content could be transitioned into a shared service model without disrupting the current accessioning model.
Potential disruption to the current ReCAP business model is an important consideration in assessing the costs and benefits of a transition to a shared print service model. However, it is equally important to consider where changes in current operations might significantly improve ReCAP’s ability to serve as a shared print service provider. For instance, a simplification and normalization of circulation rules associated with the seventy-plus customer codes would enable external clients to more easily evaluate the business value of a partnership with ReCAP. A systematic effort to consolidate ReCAP holdings under a few common access regimes would maximize the business value of the repository and likely result in more cost-effective management of the pooled resource. One obvious benefit of a cooperative collection management regime would be the space savings obtained by de-duplication of holdings transferred to the shared facility. Beyond extending the useful life of the current facility, a selective de-duplication effort would also increase the potential scope of a shared print service offer by increasing the range of titles represented in the resource.

Our study identified more than twenty-five thousand Hathi titles deposited in ReCAP by both Columbia University Libraries and NYPL. Using a cost estimate of $.86 per volume for management in a high-density facility, one can estimate that the ReCAP consortium is investing at least $40,000 annually in the management of print inventory that is duplicated within both the shared repository (with multiple partner copies on deposit) and the Hathi digital preservation repository. The long-term cost of preserving these
resources could be reduced by half or more if duplicate copies were removed from the 
repository or, better still, not accessioned into the repository. This is a conservative 
estimate, based on the number of monographic titles duplicated in both ReCAP and Hathi for 
which duplicate deposits by NYPL and Columbia could be identified. The actual cost figure is 
likely to be much greater, since in some cases multiple copies of a title have been deposited 
by each partner library. For example, Columbia University may deposit several copies of the 
same book, each under a different departmental customer code. We identified tens of 
thousands of titles in ReCAP deposited by multiple departmental libraries at Columbia 
University, which are also replicated in the HathiTrust Digital Library.

There are undoubtedly instances where duplication in the aggregate ReCAP collection is 
justified; for instance, when a title is rare or of special cultural or institutional significance. 
Our findings suggest that many titles duplicated within the ReCAP repository (i.e. deposited 
under multiple customer codes) are also held by hundreds of other libraries. As might be 
expected, titles deposited by multiple ReCAP partners are generally more widely held in the 
system-wide library collection than titles deposited by a single ReCAP partner. About a third 
of Hathi titles in ReCAP that are on deposit by a single library partner could be described as 
relatively widely-held titles, with more than 99 library holdings in the WorldCat database; 
nearly 50% of the titles deposited by multiple partner libraries fall into this “widely-held” 
category. Under the present arrangement, the collective investment made by ReCAP to 
manage these duplicate copies represents a significant opportunity cost: every dollar spent to 
store a second or third copy of a widely-held book is a dollar that can’t be spent on a 
potentially higher priority item.

Strictly speaking, it may not be possible for ReCAP to reduce significantly its expenditure on 
managing redundant inventory already accessioned in the repository. Given the effort and 
expense required to de-duplicate and re-stock inventory in a high-density repository, it seems 
unlikely that a retrospective de-selection of ReCAP holdings will be undertaken. The 
consortium could, however, **maximize the value of its ongoing investment in the 
repository collection by making it available to external library partners as a shared 
preservation resource**, analogous in some respects to the shared Hathi digital repository. By 
deliberately accessioning materials for which a broad market for service exists, and by 
managing the pooled inventory as a cooperative resource, ReCAP partner libraries can 
substantially increase the business value of the shared repository.

**Distribution of system-wide print holdings**

Finally, to understand the potential value that a shared service offering from ReCAP might 
deliver, it is useful to consider the market it would likely serve. Since libraries that hold local 
copies of titles duplicated in ReCAP and Hathi have a shared interest in the long-term
preservation of these resources and a collective interest in reducing unnecessary expenditure, we can look to the distribution of library holdings in ReCAP’s prospective service collection for an indication of its potential market for service.

![Figure 15. System-wide distribution of library holdings for Hathi titles in ReCAP (June 2010)](image)

A relatively small proportion of the titles duplicated in ReCAP and Hathi represent rare or scarcely-held resources. Less than 10% of the titles we examined are held by fewer than 10 libraries. From a shared print service perspective, this is a positive finding since we can predict that relatively few libraries are prepared to allocate significant resources to ensure print preservation of these materials, unless they can be shown to have special cultural or scholarly importance. In a shared repository environment, the costs of preserving and providing access to these resources will be comparatively low, but the value they represent to potential consumer libraries is also relatively small. By contrast, titles that are held by a larger number of libraries are likely to have a greater business value; every library that acquired this content has a greater or lesser stake in its long-term preservation and many (if not most) of them will be motivated to reassess the costs and benefits of local management once the title is represented in the mass-digitized corpus.

Based on our analysis, the ReCAP repository holds more than half a million mass-digitized titles that are also held by 25 or more libraries. This inventory could have considerable
business value as a shared print service collection. Compared to the distribution of library holdings for titles in Hathi, the ReCAP profile shows a greater concentration of widely-held books, with 70% of titles held by more than 25 libraries. This is not an exceptional finding (there is a greater probability that ReCAP will hold a title that is relatively abundant in the larger library system versus a title that is rare) but is an important one. It suggests that ReCAP could establish a market as a shared print provider for a substantial number of libraries. This has obvious implications for future business planning.

Model Consumer Profile: NYU

Analysis of the duplication between NYU library holdings and the Hathi repository has confirmed a hypothesis framed at the outset of this project: the emerging corpus of digitized books represents a potentially viable surrogate for a substantial proportion of print book collections in academic libraries, if adequately “backed up” or reinforced by a shared print access and preservation strategy. In June 2009, approximately 20% of titles in NYU’s Bobst library (as measured by holdings in WorldCat) were duplicated in the Hathi repository; by June 2010, the rate of duplication had increased to about 30%. It is tantalizing to consider the space recovery and cost avoidance that might be achieved if the library could outsource preservation and access services for at least some of these titles to shared print and digital repositories.

In absolute numbers, the overlap in titles held by NYU and Hathi is significant. Our June 2010 analysis identified more than 700,000 titles (unique editions or manifestations) that were held in both repositories, i.e., archived in digital format by the HathiTrust and held in a tangible (usually print) format by NYU Libraries. This constitutes almost a third of the Bobst library collection, on a per title basis. Based on standard volume-equivalent measures, it represents approximately 44,000 linear feet of standard library shelving or about 55,000 assignable square feet (ASF) that could be repurposed for new uses. The chart below documents the growth in duplication between NYU and Hathi holdings over the twelve months of our study.
In addition to space recovery, the potential cost avoidance that might be achieved if redundant inventory were permanently removed from the library is considerable. Using a recently published cost model produced by economist Paul Courant, one could calculate the total annual cost savings that might be achieved by replacing 700,000 locally-managed print volumes with a surrogate service provision to be as much as $3 million per year—assuming (somewhat improbably) that all staffing, operational and facilities expenditures were adjusted to reflect the change in collection size (Courant and Nielson, 2010). Yet if even a small fraction of this cost avoidance could be achieved, it is obvious that the library might substantially reduce or, more strategically, redirect its draw on the financial resources of the university.

In reality, of course, it is unlikely that NYU or any other research institution would view duplication of even very low-use local physical holdings with titles in the Hathi archive as sufficient justification for permanent withdrawal. As noted above, an overwhelming majority of titles in the digital archive are still in copyright and therefore subject to restrictions in online availability; NYU cannot simply “replace” access to locally-held physical inventory with a link to a free digital edition. Online access to these titles will ultimately require a subscription to Google Books or another licensed aggregation. More importantly, local faculty—especially humanities scholars—will almost certainly expect NYU to provide ongoing access to print versions of titles that were purchased by the library, irrespective of online...
availability. This leads us to the question of whether a regional storage collection like ReCAP can provide a more cost-effective solution to long term physical preservation and access, for titles that have been digitized and securely archived by Hathi.

The level of duplication between NYU, Hathi and the ReCAP facility provides a baseline measure of the potential savings that might be achieved in the near term.

![Figure 17. NYU Bobst titles duplicated in ReCAP and HathiTrust Digital Library (September 2009 - June 2010)](image)

A relatively small proportion of the titles owned by NYU that are currently replicated in the HathiTrust Digital Library were found to be duplicated in the ReCAP sample data. Based on our June 2010 analysis, we identified about 200,000 titles that might be eligible for withdrawal from NYU's Bobst Library collection, based on “dual duplication” in Hathi and ReCAP, if a robust shared service agreement were in place. This figure represents about 10% of the Bobst Library’s total holdings in WorldCat, a relatively modest figure compared to the 30% of Bobst holdings that are replicated in the mass-digitized corpus in Hathi. If we limit the analysis to monographic titles in the public domain, the total number of books that would be considered in scope for a shared print agreement between NYU and ReCAP dwindles sharply, to about 18,000. In short, the initial premise that a cooperative management regime restricted to books in the public domain might deliver sufficient benefit
to mobilize a significant change in local library operations is seemingly not borne out by our findings.

If we look instead at the full range of Bobst holdings in Hathi—including in-copyright titles—ReCAP appears to be a potentially more valuable supplier of print preservation and access services. More than 90% of the 200,000 titles in Bobst that are duplicated in Hathi represent in-copyright content; nearly all of them (93%) are designated as unrestricted holdings in the ReCAP collections. A print supply chain for these titles will remain indispensible for some time to come and whether post-digitization demand increases or falls, centralized repository services will prove more cost-effective solution than local management. Accordingly, one might concede that a shared print service offer from ReCAP that includes any mass-digitized title owned by NYU would maximize the library’s ability to reduce unnecessary expenditure and redeploy resources in support of services more directly tied to the university’s research and teaching mission.

Even so, it is not clear that a shared print service arrangement with a sole provider like ReCAP can deliver the level of benefit that is ultimately desired. Is a shared print offer that enables a net space recovery of less than 13,000 linear feet of shelving (or approximately 16,000 ASF) in its first year worth the energy and effort that would be needed to draft and implement a binding agreement? At the time this study began, NYU was in the midst of a major library renovation project that required the removal of nearly 500,000 volumes from the Bobst library building. Under normal operating circumstances, the library routinely transfers 100,000 volumes or more to storage each year, simply to accommodate annual growth in the print collection. By comparison, the potential space savings and cost avoidance achieved by outsourcing collection services for 200,000 books in ReCAP appears relatively small.

Realistically, one can estimate that a potential consumer of shared print services might hope to “externalize” or outsource management for all, or almost all, of the mass-digitized content in the local collection. Thus, NYU might reasonably seek to negotiate a shared service agreement that would cover most of the 700,000 locally-owned titles that are currently duplicated in the HathiTrust Digital Library, while also allowing for additional growth in years to come. Under present conditions, a shared print agreement with a single supplier with ReCAP’s current collection profile would deliver less than a third of the value that NYU might hope to derive from a fully satisfactory shared print agreement that provided comprehensive coverage of the mass-digitized corpus.

Does this mean that academic libraries must postpone any space and cost saving reorganization of low-use print collections until an eventual e-licensing option for the mass-digitized book corpus emerges? We think not. A range of options are available to individual
academic institutions seeking to externalize some part of the operational and cost burdens associated with managing a low or no-use legacy print collection. Several of these options are explored below.
Shared Print Provision: Assessing the Options

As outlined above, our analysis of the NYU library holdings duplicated in the HathiTrust Digital Library and ReCAP suggests that the total space savings and cost avoidance that is achievable through a shared print agreement as originally conceived is relatively limited at present. This outcome is dependent on a number of variables, at least some of which are subject to library influence or control. Libraries seriously motivated to design and implement shared service agreements will want to consider all available options for maximizing the impact and sustainability of these models.

Expanding the Scope of Shared Service

In examining the NYU library collection, we limited our study to holdings in the Elmer Holmes Bobst library, which serves the general undergraduate population as well as graduate researchers and faculty in the humanities, social and natural sciences. Bobst is the largest of NYU’s libraries, with holdings in excess of 5 million volumes, and ranks among the top academic research libraries in the United States. Our analysis found only a modest overlap between holdings in Bobst, Hathi and ReCAP, amounting to less than 10% of all titles in the Bobst library collection. By contrast, the duplication rate between Bobst and Hathi alone is estimated to exceed 30% of the local collection.

Based on library holdings registered in the WorldCat database, we estimate that holdings in Bobst account for about 85% of all titles held by NYU libraries. Conceivably, a greater yield might be obtained in a shared print agreement with ReCAP if the scope of our analysis were adjusted to include a wider range of NYU library units. To test this hypothesis, we expanded the scope of comparison to include all of the NYU libraries with holdings set in WorldCat and identified a total of 775,980 unique titles replicated in Hathi as of June 2010. This represents a 10% increase over the number of Bobst titles in Hathi (679,401 titles). Yet while the base of comparison for a shared print offer for NYU was larger in absolute terms, the proportion of titles that ReCAP might supply actually decreased to 28%, compared to 29% for Bobst alone. In a very real sense, one can say that expanding the scope of an initial shared print service offer to include a broader range of library types (including specialized departmental libraries)
would actually result in a lower net benefit. Expanding the scope of a shared print arrangement would also entail more complex business agreement since not all NYU libraries are under the same administrative or budgetary control.

This leads us to conclude that an initial shared print offer should in the first instance be scoped around the space- and cost-saving objectives of a limited range of academic libraries that share a common set of service expectations. Based on what we have seen from the shared digital and shared print repository profiles, the target audience is likely to be moderate to very-large college and university collections in the humanities and social sciences. *One can predict that mid-size universities with a strong commitment to the humanities, along with liberal arts colleges, will represent the core market for shared monographic preservation and access services, since they are committed to uphold a preservation mandate for which local resources are increasingly inadequate.*

**Assessing Market Maturity**

The findings reported here are based on a twelve-month study of the mass-digitized corpus in Hathi and a single, partial snapshot of the ReCAP print repository. As shown in Figures 10 and 17 above, the prospective value of ReCAP as a shared print service provider has increased significantly in the past year based on the rapidly expanding scope of Hathi alone. During this same period, the ReCAP collection itself has also grown, with about 50,000 new items representing an unknown number of unique titles accessioned each month on average. Is it possible that the “dual duplication” rate that was predicated as necessary for shared service provision—with at least one copy in Hathi and one copy in a shared print preservation repository—will increase as the ReCAP inventory continues to grow? If so, is it worth waiting until the duplication between ReCAP and Hathi reaches a desired threshold? This question is especially pertinent in the case of ReCAP since all three ReCAP partner libraries have joined the HathiTrust since the inception of this project. These libraries are now more likely to transfer to ReCAP the titles digitized from their local collections, which will naturally increase the match rate between Hathi and ReCAP.

The evidence in hand suggests that unless a deliberate and systematic effort is made to align shared print repository holdings and Hathi digital repository holdings, it is unlikely that the existing preservation infrastructure embodied in large library storage collections will coalesce into a sufficiently robust source of surrogate supply. What is required is not an incremental and *ad hoc* change in storage transfer protocols at individual repositories, but a purposeful and coordinated strategy to create a shared print infrastructure capable of delivering significant tangible benefit to a large number of academic libraries. Deferring the negotiation of shared print agreements until such time existing repositories exhibit the desired service profile will simply delay the development of shared infrastructure. Instead, *library*
administrators who readily perceive and are prepared to realize the benefits of selectively outsourcing print management functions to a shared service provider can accelerate the process by stipulating clear expectations and establishing targets that prospective providers will be motivated to meet.

Alternative Service Providers

We have seen from our case study that the current ReCAP repository collection, as represented in our sample, is not optimized for shared print provision as originally envisioned. At best, it might at the outset offer a client such as NYU surrogate preservation and access services for about a fifth of the mass-digitized book collection in Hathi and enable a local space recovery of about 13,000 linear feet, or the equivalent of about 200,000 volumes. Even if this figure were to grow in future years, as will almost certainly happen, the initial value proposition appears insufficient to justify a service agreement, except perhaps as a symbolic gesture.

Could another large scale print repository provide a more competitive offer? The Southern Regional Library Facility (SRLF), a shared library storage facility serving five campuses in the University of California (UC) system, currently holds about 6 million items, a collection about three-quarters the size of ReCAP. The SRLF provides a useful counterpoint to ReCAP, since it represents holdings from a more uniform base of academic institutions and encompasses a broader range of university libraries, including both ARL and non-ARL institutions. Inventory at the SRLF is managed as a cooperative resource with a non-duplication policy applied across the aggregate collection. Thus, one might expect that the SRLF collection would be more broadly representative of academic library holdings and also contain a greater proportion of unique titles than is the case at ReCAP. The SRLF is remarkable in another way as well: it is one of the few large library storage collections whose holdings are represented in the WorldCat database under a discrete library symbol.
Although smaller in absolute size, the SRLF appears to offer greater initial value as a shared print partner for NYU. Comparing Figures 17 and 18, one can see that the SRLF would enable a marginally greater space savings and potential cost avoidance than ReCAP. More significantly, perhaps, it appears that the rate of growth in coverage over time is sustained and relatively stable—important factors when negotiating a business agreement that is in part predicated on a forecast of future value. Additionally, a visible spike in the proportion of public domain titles that the SRLF could supply between May and June 2010 serves as a useful reminder that a shared print repository that proactively contributes to the development of a shared digital infrastructure simultaneously increases its value as provider of surrogate print services. The increase in public domain titles held in common by SRLF and NYU in June 2010 reflects a large Hathi ingest of titles digitized by the University of California in partnership with Microsoft and the Internet Archive.

The UC SRLF example is instructive, but not necessarily representative of the broader marketplace of potential shared print providers. Over the course of this project, another large library preservation repository fortuitously became visible in WorldCat: the UC Northern Regional Library Facility (NRLF), which serves five campuses in Northern California. While approximately the same size as the SRLF, with approximately 5.5 million items held in June 2010, the NRLF offers significantly less coverage of the mass-digitized book corpus held by NYU, though still more than ReCAP can presently provide. This is due in part to the fact that

---

**Figure 18. NYU Bobst titles duplicated in UC SRLF and HathiTrust Digital Library (June 2009 - June 2010)**

![Figure 18](chart.png)
the NRLF holds a greater proportion of rare and unique titles than its Southern California counterpart; as a result, there is a lower probability that titles in the repository will be duplicated in Hathi (unless directly contributed as digitized content) or in other library print collections. As noted above, repositories that hold a greater relative proportion of titles with very low aggregate holdings will probably find it difficult to establish significant market share in a shared print service environment.

Compared to ReCAP, the University of California's two massive regional repositories offer only slightly greater coverage: the UC NRLF might provide a provisioning option for 30% of the titles of interest (compared to 29% at ReCAP); the UC SRLF could potentially provide preservation and access services for 36% of the assumed target of 700,000 volumes.

![Figure 19. Comparison of potential shared print provision options for NYU Bobst Library (June 2010)](image)

Even if one were prepared to accept the additional challenges of ensuring timely on-demand physical delivery from one side of the continent to the other—which might reasonably be addressed in a shared service business model—it is difficult to imagine that a marginal gain of fifty thousand titles or a few thousand linear feet of shelf space would motivate a library like NYU to go farther afield for a shared print partner.

A potentially more profitable alternative would be to contract with multiple shared print repositories to increase the scope of coverage and maximize the externalization of high-cost,
low-return print operations. Yet, as shown in Figure 19. Comparison of Potential Shared Print Provision Options for NYU Bobst Library, even the combination of multiple large scale repositories results in only a relatively modest increase in the tangible value of the partnership: the union of three very large shared print collections scarcely suffices to replace 60% of the mass-digitized titles in NYU’s Bobst Library. A recent study (Payne, 2007) identified approximately 70 high-density academic library storage facilities in North America; each one of these repositories may be said to represent a potential shared print service provider. Multi-lateral agreements could theoretically be struck across this network of repositories to maximize the scope of shared print agreements. At present it is virtually impossible to assess the carrying capacity of this infrastructure, since only a small number of high-density storage collections are currently visible in national and international union catalogs. Until the latent value of this aggregate resource is effectively and systematically disclosed, it will be difficult for individual libraries to judge the benefit of prospective shared print partnerships.

**Optimizing Existing Infrastructure**

Based on our necessarily limited view of existing infrastructure, there is a significant gap between the level of shared service provision a client library like NYU might reasonably seek and what repositories like ReCAP or RLF might readily provide. This does not reflect an intrinsic flaw in the library system; it is the necessary and natural outcome of a business model in which print collections are acquired and managed primarily as a local resource. In most instances, withdrawal and storage transfer decisions are reactive responses to local space crises and not intentionally guided by long-term library strategy. As a result, off-site depository collections as currently constituted have only limited value as a source of surrogate print preservation and access services for “external” consumers. Their business value as a cooperative resource is correspondingly small.

Ultimately, the benefit and business value that shared print repositories can deliver will be determined not by present inventory or service capacity, but by their individual and collective ambition to transform the academic library enterprise. If the market for shared service is sufficiently great, even commercial interests may be motivated to acquire and manage print inventory on behalf of academic libraries. More probably, some number of existing library repositories will opt to reconfigure collections and operations to support shared service provision, so that academic institutions can outsource management of low use print holdings. This would effectively result in an optimization of the existing library infrastructure as resources are pooled to meet collective service requirements.
We can judge the potential impact of this reorganization by returning to the now familiar NYU use case. If we expand the scope of analysis beyond ReCAP itself (as represented in our sample) to include the totality of holdings in the ReCAP partner libraries—on the presumption that any title held by Columbia, Princeton or NYPL might eventually be transferred to the shared repository—we find that NYU could potentially outsource print management of more than 90% of the mass-digitized titles in its collection. Moreover, comparing Figures 16 and 20, one can see that over the twelve months of our study, the ReCAP libraries were consistently capable of supplying more than 90% of the mass-digitized titles in NYU’s collection. Simply put, ReCAP has the potential to satisfy NYU’s anticipated shared print service need.

![Figure 20. NYU Bobst titles duplicated in ReCAP partner libraries and HathiTrust Digital Library (June 2009 - June 2010)](image)

One might conclude from this that NYU would be best served by striking a shared service agreement with the ReCAP partner libraries directly: why wait for the desired inventory to be transferred to the storage facility?

Practically speaking, it would be difficult for NYU to negotiate with the ReCAP consortium for guaranteed access to print collections still managed on-site at Columbia, Princeton and NYPL. The ReCAP consortium was formed to provide a governance structure for the shared library storage facility and does not exercise any joint authority over the partner library collections. NYU would instead need to negotiate individual agreements with each of the partner libraries. Beyond the administrative overhead of negotiating and implementing multiple agreements, there are other factors to consider. First, collections managed on-site have a
higher loss rate than collections managed in a high-density repository and are subject to variable environmental control; the preservation value of individually negotiated agreements will therefore be less than ReCAP can provide. Secondly, the direct cost of managing on-site collections is substantially greater than managing collections in storage, so the shared access agreements would likely be more expensive as well. Finally and most importantly, ReCAP partner collections are not managed as a cooperative resource and as a consequence NYU could not obtain any meaningful assurance that the shared collection(s) would be subject to similar terms and conditions.

As an alternative, NYU could conceivably negotiate a shared print agreement with the ReCAP consortium members that is contingent on the prospective transfer of materials to the shared storage facility. In addition to securing NYU guaranteed access to these materials under common terms and conditions, an agreement like this would deliver benefit to the consortium members by creating an incentive to implement a genuine cooperative management plan that would hasten the transfer of materials to storage (reducing inventory management costs for individual contributors) and potentially reduce or eliminate duplicate deposits (extending the useful life of the facility). Moreover, inasmuch as titles held in common by NYU and ReCAP members are also likely to be held by other academic libraries in the region, accelerated transfer of these materials to ReCAP would increase the likelihood of a scalable shared print business model.

Striking an agreement of this kind, which explicitly articulates expectations of prospective and targeted growth in a shared print service collection, and also leverages the full value of repository infrastructure, would undoubtedly be challenging. The initial gains for NYU and for ReCAP might be relatively modest, but the long-term benefits could be genuinely transformative. NYU, Columbia, Princeton and NYPL could all reasonably expect to regain significant library space and reduce redundant expenditure while contributing to the creation of an optimized and sustainable print collection. Such an agreement might also serve as a model that could be adapted for use with other large-scale storage repositories, expanding the total market for shared print service and helping to establish common norms and expectations. For these reasons alone, a valiant effort is arguably justified.
What is It Worth? Putting a Price on Shared Collection Services

Thus far we have characterized the value of shared collection services primarily in terms of library space recovery: linear feet of shelving that might be freed up to accommodate new acquisitions or assignable square feet that might be repurposed as study space, learning and research commons, etc. However, a deliberate strategy to externalize collection management functions for low use print inventory can also generate economic benefit in the form of cost avoidance for consumer libraries and cost recovery for shared service providers. Even more importantly, joint business agreements that effectively redistribute the costs and benefits of print and digital preservation have the capacity to transform the academic library enterprise, freeing individual organizations to pursue service goals that are relevant to local needs while still meeting collective stewardship obligations.

In real terms, we can estimate the economic value that might be obtained through an externalization of print management functions with a simple (and admittedly oversimplified) cost calculation. For a potential consumer such as NYU, the total cost avoidance will be determined by the number of volumes that are covered by a shared print service agreement. Based on the analysis described above, we judge that under present conditions ReCAP could potentially provide surrogate collection services for about 200,000 mass-digitized titles in NYU’s Bobst library. Since nearly all of the titles are monographic publications, we will conservatively estimate that each title represents a single volume in the Bobst collection. Using an estimate of $4.26 per volume, we calculate the cost to manage these titles on-site in the Bobst library to be approximately $850,000 per year. In an efficient high-density storage environment like ReCAP, the annual cost of keeping the same number of volumes is significantly less: about $172,000 based an estimate of $0.86 per volume. By outsourcing inventory management functions to ReCAP, NYU might therefore achieve a significant reduction in the per-unit cost of preserving these books, in addition to regaining about 13,000 linear feet of library shelving or 16,000 ASF of space in Bobst.

Of course, since the cost estimate for keeping a book includes a number of “sunk” costs, NYU will not actually recover an amount equal to—or remotely approaching—$850,000 each year.
The retail value of the ReCAP service offer is more likely to approximate the savings NYU will realize by not transferring these titles to its own storage facility. Because NYU leases a facility and has thereby avoided the up-front capital costs of construction, the per-volume lifecycle management costs are almost certainly less than $0.86 per year. It stands to reason that NYU would therefore reject a shared print service proposal for which the cost exceeds $170,000 per year. Normal ReCAP operating expenses are already subsidized by the three consortium members, so the marginal costs of offering service to NYU and other libraries would likely serve as the basis of a pricing model, allowing for future growth in the scope of the service collection and projected opportunity for space savings at the client libraries. An ambitious shared print service provider could maximize cost recovery by building a service collection shaped to the needs of an external clientele, substantially reducing or even eliminating the charge backs customarily used to sustain shared storage repositories. For privately funded organizations like ReCAP, as well as publicly funded entities like the UC Regional Library Facilities, the external market for shared print service may represent a path to long-term sustainability.

This is not the place for a comprehensive examination of business models that might support shared print service; logically, that work will be taken up by organizations that aspire to serve as service providers, in consultation with motivated consumers. Instead, we can offer a few tentative observations based on findings from our empirical study of existing infrastructure and anticipated service requirements. If shared service provision is to be developed on the backbone of the existing storage repository infrastructure, as seems likely in the near term, it will be necessary to strike a balance between the need to minimize retrievals from high-density facilities, which would tend toward an “insurance only” access model, and the interest in maximizing reliance on external providers, which would tend to concentrate demand on a small number of suppliers. Both of these goals could be accommodated in an arrangement in which pricing is determined in part by the demand profile of the service collection.

Returning to the example of NYU, we can anticipate that demand for the 200,000 titles that might be covered in an initial agreement with ReCAP, will vary according to more or less predictable patterns. A pricing scheme that is sensitive to this variability would protect ReCAP from the negative cost consequences of increased retrievals while providing NYU the guaranteed access it requires. Since overall demand is likely to be low across the service collection, a transaction-based pricing model seems inadvisable; ReCAP or any other large library storage repository would find it difficult to generate a reliable stream of cost recovery even if shared access agreements were struck with a large number of client libraries in the region. Instead, an annual baseline contribution from client libraries, pro-rated according to the size and value of the service collection, might offset normal operating expenses, while a
variable fee based on the specific demand profile of the titles covered by the agreement would provide necessary flexibility.

Aggregate demand patterns—for example, the higher circulation rates that are typically observed (among North American libraries) for English-language publications of relatively recent vintage, or the extremely low use profile of monographs with non-roman scripts—could be used to establish the overall demand profile of a service collection. This would allow for standardization of demand-based pricing, which would improve the transparency of shared print business agreements. This in turn might stimulate healthy competition amongst shared print service providers. If well-publicized aggregate demand patterns were used to establish pricing rates, individual shared print repositories could establish a competitive advantage by offering comparable service at varying price points.

Discriminatory pricing based on demand profile would allow shared print providers to tailor service agreements according to the needs of institutional subscribers. In the case of NYU, more than half of the titles for which a shared service agreement might presently be struck with ReCAP represent English language monographs published in the last decade. The higher demand profile for these titles compared other ReCAP holdings would reasonably justify a higher service cost. In theory, pricing and service level agreements might be exquisitely sensitive to variations in demand; practically speaking, a simpler model is likely to prevail, if only because library organizations are not especially entrepreneurial in nature. Ideally, a demand-based pricing model would allow for periodic adjustments based on changes in the library system as a whole, so that the market for shared print service is not subject to dramatic fluctuation, which might have devastating consequences for individual libraries that find themselves priced out of the local marketplace.

From a consumer perspective, the operational value of a shared print agreement offer will be determined by the initial scope and size of the service collection, and its rate of growth over time. Libraries contracting for shared print services will want to achieve maximum benefit in the form of local space recovery and cost avoidance, which is dependent on the scope of the service collection. They will also reasonably seek an assurance of continued growth in the service collection, since a library’s ability to derive ongoing benefit from the arrangement requires that new space savings be gained each year. From a service provider perspective, the costs of delivering shared collection service will be determined by the rate at which material destined for a service collection is accessioned and the rate at which it must be supplied. It is in the mutual interest of shared service providers and consumers that repository collections rapidly assume the profile of “optimal” service collections, so one can anticipate that prospective providers will as a matter of course begin to accession inventory according to market needs.
It is worth considering that the increased discoverability of the mass-digitized book collection may result in greater demand for the print version, especially in the absence of a licensing agreement for the in-copyright titles. A recent study of post-digitization use of print collections at the University of Michigan found that the increased discoverability of books made available as full-text resources online did not result in increased demand for locally-held print versions (Look, 2010). The scope of the study was small and only addressed titles in the public domain so it is not possible to infer that demand for the much larger in-copyright corpus will be similarly unaffected by increased network visibility. Further analysis of aggregate demand patterns for titles already in storage could provide useful insights into the likely impact of pooling supply and demand for low- and moderate-use academic print collections on a regional basis.

Because retrievals are the single greatest cost driver in high-density facilities, repository managers are motivated to control demand for physical inventory by accessioning only (or mostly) low-use titles. In a shared service context, there is some risk that the concentration of demand from multiple institutions will result in increased retrievals and higher operating costs. This risk could be mitigated by aligning shared print service collections with the mass-digitized book corpus in Hathi and ensuring that digital surrogates are the primary mode of discovery and delivery. This alignment would serve a dual purpose by maximizing the benefit libraries can derive from the mass-digitization enterprise while also providing a means of moderating physical retrieval rates. The result would be a virtuous circle of shared service provision, in which collective library investment in the creation of the HathiTrust Digital Library is repaid by the increased efficiency in library operations enabled by cooperative print management.

**Who Will Benefit? Who Will Pay?**

We have established that a deliberate reorganization of the existing ReCAP collection in which inventory is more closely aligned with the growing corpus of mass-digitized texts, along with other dual format titles, would substantially improve its ability to function as a shared print service provider for NYU. We have also examined the distribution of library holdings in both Hathi and ReCAP and hypothesized that there is a substantial market for shared service based on the many hundreds of thousands of titles for which aggregate library holdings are relatively abundant and demand is low, and for which a shared service provision based on existing repository holdings appears feasible. Is the potential market for service sufficiently large to sustain shared print service at scale? Can a core segment be identified for which a common model of service provision might be satisfactory?

To answer these questions, we returned to the constituency from which this project was born: university-based academic research libraries in North America. Measuring the
percentage duplication of titles in each of the 113 ARL university libraries and the HathiTrust Digital Library at twelve-month intervals, we established a baseline against which our findings for NYU could be compared. As shown in Figure 21, the results indicate remarkably low variance in duplication levels across the ARL cohort. This is an especially notable finding since there are great disparities in the library volume (and respective title) counts among ARL libraries, ranging from more than 16 million volumes at Harvard University to fewer than 2 million volumes at the University of Guelph, based on data reported to ARL in 2007-2008. In June 2009, an average of 20% of titles held in any given ARL library was duplicated in the HathiTrust Digital Library; by June 2010, the average duplication rate had increased to 30%. These figures are consistent with the levels we found for NYU’s Bobst Library.

![Figure 21. Percentage duplication of titles held in ARL libraries and HathiTrust Digital Library (June 2009 and June 2010)](image)

This scatter chart provides a simple but effective visualization of an important pattern that this project has revealed: that is, that the risks and opportunities associated with moving collection management “into the cloud” are uniformly distributed across the research library community as a whole.

Based on these findings, we estimate that the median space savings that could be achieved at an ARL library if a robust shared print offer were in place today to be approximately **36,000 linear feet or the equivalent of more than 45,000 ASF**. To put this in perspective it is helpful to consider that the space requirement for a typical library-based research or learning...
commons is about 20,000 square feet. In economic terms, the total annual cost avoidance—assuming all of these books are currently managed on-site—exceeds $2 million per library. As noted above, this is not a sound basis for judging how much library resource would actually be available for redirection in support of other operations; however, it does provide a useful measure of the opportunity costs of inaction. The cost of managing these same titles in a high-density off-site facility would amount to approximately $500,000 per year for each library. This provides a rough measure of the maximum retail value for a shared print offer that would enable a full externalization of print operations for mass-digitized books.

These figures represent a necessary oversimplification of what is obviously a much more complex and challenging business case. For example, a significant number of ARL libraries own off-site shelving facilities and it is likely that some of the mass-digitized titles that might otherwise be relegated in favor of a shared service agreement are already “locked up” in storage. The calculation of benefit for those libraries would necessarily be different. By the same token, some libraries that have already transferred a significant portion of the locally owned inventory to storage may see themselves as potential shared print providers, rather than consumers. It is not possible to provide an accurate forecast of which libraries will likely assume a role as a supplier or consumer; nor it useful to speculate about those that will disdain the shared service model, preferring instead to operate in relative isolation. Still, we feel that our broad brush estimate of the market for shared print service in the ARL community provides a useful starting point for library directors and repository managers to begin planning and even budgeting for a future in which managing local print collections is no longer a core function or cost center.
Conclusions and Recommendations

Our year-long study of the mass-digitized book corpus in the HathiTrust Digital Library and parallel investigation of potential shared print service providers has confirmed that there is an opportunity for significant library space savings and cost avoidance if management operations for digitized books are deliberately and systematically outsourced or externalized to shared service providers.

One can anticipate that academic institutions interested in reducing local print holdings in favor of regionally consolidated inventory will, in years to come, increasingly look to extant repositories like ReCAP, the UC Regional Library Facilities, etc. as a source of preservation and access services. Our findings suggest that current storage inventory is not presently optimized to support shared print solutions on a large scale, but also indicates that system-wide reorganization of collections and services that maximizes the business value of print as a cooperative resource is both feasible and capable of producing great benefit to the academic library community.

Our findings suggest that the shared infrastructure that is needed to support a broad-based externalization of legacy print management functions is unlikely to materialize without some purposeful action by the academic library community. By describing and—where possible—quantifying the value that a changed infrastructure might deliver, we hope to have contributed in some measure to stimulating potential consumers and suppliers alike. Further work will be needed before academic libraries and the educational institutions they serve can fully realize the benefits of shared service:

- It is in the interest of all academic libraries that mass-digitized collections be made more widely available to students and researchers, and that their scope and quality improve to the degree that low-use print inventory can be retired in favor increased reliance on digital surrogate. Library directors and academic administrators should advocate in favor of licensed access to the mass-digitized resource as part of a comprehensive strategic plan in which the library can reassert its role as a vital part of the academic enterprise.
The HathiTrust’s ongoing efforts to expand public access to the mass-digitized book corpus through programmatic rights assessment, direct negotiation with rights holders, and by accessioning large aggregations of digitized public domain resources, should be recognized as a major contribution to the transformation of the library service environment. By developing a partnership model that is not dependent on content contribution, Hathi can deliver benefit to an even broader range of academic institutions.

- Beginning in 2013, the HathiTrust will introduce a new cost model that will enable non-content contributing members to participate in (and benefit from) joint stewardship of the digital repository. This new affiliation model is described in Appendix I.

Institutions and organizations that aspire to roles as shared print service providers need to proactively build collections that will deliver maximum operational value to external audiences; they should leverage the collective library investment in Hathi by accelerating the transfer of mass-digitized titles to print preservation repositories and self-consciously promote these resources as shared and cooperatively managed assets. Above all, these institutions must take steps to make their service ambitions and capacity known, so that potential consumers can begin to articulate core requirements and common service profiles can be identified.

- In cooperation with other library organizations, OCLC is working to develop technical solutions that will enable the latent value of library storage collections and distributed print archives to be more effectively disclosed.

Prospective shared print providers can help to define a common service profile by surfacing model agreements and engaging in dialog about the operational and business requirements of shared service provision. Managers of large print repository collections should be empowered and encouraged to engage in business modeling exercises that are explicitly intended to expand the market for service beyond content contributors.

- ReCAP partner libraries have outlined core elements of a shared service agreement. These are summarized in Appendix II.

Libraries that are already motivated to outsource legacy print management functions in support of a changed service portfolio or simply to relieve local space pressures should begin to establish objective targets, and quantify and articulate desired outcomes so that motivated suppliers can respond in kind. Library administrators should engage directly with faculty and academic officers to communicate a compelling strategy in which selective externalization of traditional functions is
demonstrably improving the institution’s ability to fulfill an academic and research mission. This work will be challenging and deserves external support and endorsement by library leadership organizations and funders.

• Research organizations can advance our collective understanding of the changing profile of demand for legacy print in the mass-digitized environment and help to characterize the optimal redistribution of library resources.

When these steps are taken, we will have made measurable progress toward the worthy goal of ensuring the long-term survivability of the scholarly record at a cost that is sustainable for the research library community as a whole.
Appendix I. HathiTrust Cost Rationale

Beginning in 2013, HathiTrust will use a cost model that reflects benefits that partners can receive from works stored in HathiTrust rather than the cost associated with storing them. We believe that this new cost model, focusing on consumer values rather than storage costs, better reflects the long-term interests of partners and will more fairly distribute costs across the partners.

Beginning in 2013, HathiTrust will use a cost model that reflects benefits that partners can receive from works stored in HathiTrust rather than the cost associated with storing them. We believe that this new cost model, focusing on consumer values rather than storage costs, better reflects the long-term interests of partners and will more fairly distribute costs across the partners.

This new benefits-based cost model attributes the cost of storing a volume to each library that holds (or held)\(^1\) a corresponding print volume. In this model, a library that pays for a share of the cost of storing content is also acknowledged as receiving the benefit of the content. Those benefits for in-copyright volumes are not always tangible, but include producing replacement copies and offering specialized services permitted by contract or law. Such a cost model would, for in-copyright works, attribute a share of costs to all partnering libraries that hold or held the corresponding print volume. Because all member libraries enjoy the benefits of public domain works, every partnering library will be assumed to hold these works.

This new model of cost attribution will, compared to the current storage-based model, have a smoothing effect, reducing the cost borne by an institution that contributes significant content, and recovering cost from each member of a new class of partner libraries (“Sustaining Partner” libraries) that shares in the benefit of these volumes.\(^2\)

\(^1\) Particularly for Section 108 uses, a library may wish to withdraw (and thus no longer hold) a volume. We should store information that shows that this library once held the volume in question.

\(^2\) The philosophy of collective costs and collective holdings already underpins much of the CIC approach to HathiTrust.
The current HathiTrust membership is comprised primarily of institutions contributing large amounts of content and, thus, bearing large costs. At this time (late 2009), we are in conversation with research libraries that do not have large amounts of content to contribute, but wish to join HathiTrust to participate in its curatorial work. Under the current model, partners with smaller amounts of content pay relatively small amounts and have access to large amounts of content. By applying a model based on shared holdings, we will see some reduction in the cost per contributing institution as more libraries join the effort. Clearly, as these new libraries join, it must be with the understanding that a new cost model will work to distribute these costs in an equitable way that reflects the benefits accruing to all partners.

This holdings-based cost model will incorporate a number of precise elements about costs and the “sharedness” of the content. The cost will be re-calibrated each year, as costs for infrastructure will change each year. It is also the case that the current partners would like to be able to use shared funds to develop new services and functionality. Consequently, we will multiply the cost of infrastructure by a variable amount (adjusted periodically) to fund those new services and functionality. In this new holdings-based cost model, the costs to an institution per year will be calculated as follows:

**For public domain volumes:**
\[(PD \times X \times C)/N\]

- where
  - PD is the total number of public domain volumes in HathiTrust (assumed to be “held” by all partner libraries). This number will also include in-copyright works where the rights holder has given the members free use of the content.
  - C is the average annual cost to provide basic support for a volume. Note that costs will vary by volume, as each printed volume will vary in number of pages and in average file size. We will use an average cost that will be periodically recalculated.
  - X (greater than one) is a value with which we multiply C to generate a surplus.
  - N is the total number of partner libraries.

**For a given in copyright volume, IC:**
\[IC = (C \times X)/H\]

- where
  - C is the average annual cost to provide basic support for a volume (as above).
X (greater than one) is a value with which we multiply C to generate a surplus.

H is the number of partner libraries that hold a given print IC volume.

Initially, HathiTrust proposes using a value of two (2) for X, i.e., doubling the cost of maintenance in order to build a fund for services. Thus, 50% of the funds collected will go to development and the other 50% will cover the costs of storing content, again lowering cost proportionally for institutions that contribute large numbers of commonly held volumes.

We believe that this new model will be both equitable and sustainable, but acknowledge that it also presents significant challenges. The biggest of these is the lack of information in the library community currently, on a volume-by-volume and institution-by-institution basis, about overlap in our print collections. No organization, including OCLC, stores information about holdings, and even where OCLC succeeds in approximating this, it lacks volume-specific information. We also face challenges with regard to reliable data. For example, although each of our institutions individually stores volume-specific information, enumeration and chronology information is represented so variously that manual remediation will be required to make it uniform. Ensuring that this information is up to date is an additional concern. HathiTrust, either by itself or in collaboration with another organization, will attempt to create a system to store partner holdings information in such a way that it can be updated constantly, by partner institutions themselves or by central HathiTrust staff.

We believe that this volume-specific infrastructure will be valuable for a number of purposes, including:

- **de-duplication:** although duplication of contents is not costly, storing duplicates compromises the user experience and it obscures collection development needs;

- **management of corresponding print volumes:** how will we know that we can withdraw a print journal without having volume-specific information?

- **legal uses of in-copyright materials:** For example, Section 108 uses will depend on having a clear sense of which institutions own(ed) which print volumes.

A clear sense of volume-specific information for digital materials and corresponding print volumes will be needed as our collaboration in HathiTrust develops. HathiTrust plans to launch this new cost model in 2013, during the second phase of our initiative (i.e., subsequent to the first five years of HathiTrust). Prior to that time, we will work to develop the necessary infrastructure to be able to perform these calculations reliably. We will also, effective immediately, entertain membership from other research libraries that wish to share this curatorial role. To calculate costs for these Sustaining Partner libraries, we will use
overlap formulas developed in our current explorations of this model with RLG, ReCAP and New York University Library. All funds generated through the participation of Sustaining Partner libraries prior to 2013 will be devoted the development of the new common holdings infrastructure.

jpw, 12 Feb 2010
Appendix II. Cloud Library Service Agreements: ReCAP as Shared Print Repository

DRAFT FOR DISCUSSION

Assumptions:

The objective is to define a service agreement under which ReCAP will provide access to a defined set of print materials under specified conditions, allowing other libraries (“clients”) to discard their own print copies. Several basic assumptions inform this model, but each leaves room for some variance in interpretation.

1) A digital copy will be readily available to the client library’s patrons.

2) The agreement applies only to copies once owned by the client library and since withdrawn, i.e., the agreement is intended to allow client libraries to save the expense of storing print materials, not to expand the client’s collections, or to serve as a back-up to its print collection.

3) The agreement must provide reasonable assurance of continued access to the defined materials. It will therefore limit the freedom of action of ReCAP partners to remove materials from ReCAP permanently or place future restrictions on use.

4) The agreement will not restrict the use of the defined materials by the owning ReCAP partner’s own patrons, nor will it carry an obligation to replace items lost or damaged through normal use, i.e., the level of assurance of long-term access will be comparable to that which would be provided for the client’s own unrestricted collections.

5) The agreement must provide a level of access and service greater than that available through standard interlibrary loan.
From the client perspective, there is a further assumption: that the digital copy will be preserved and will continue to be accessible. In the Cloud Library project, the implications of this assumption are being tested through a model service agreement with the Hathi Trust. From the ReCAP perspective, the key issue is continued availability of the digital copy to the client library’s patrons. The means of assuring that availability is relevant only to the extent that it defines the scope of the agreement (discussed further below.)

Elements of an agreement and related issues:

General considerations:

Any agreement might be closely defined and tightly construed, or might be left relatively open, allowing for greater flexibility. Similarly, an agreement might place greater obligations on ReCAP or on the client, not only with regard to the service provisions, but also with regard to the steps needed for implementation. As an experiment with a new model, an agreement between ReCAP and a single client might benefit from greater flexibility, both in definition and execution. Ultimately, ReCAP would want to have the same agreement with multiple clients, and a single library might want agreements with several repositories. That would argue for closer specification of terms.

Governance:

In the Cloud Library project, ReCAP serves as a representative regional repository. In reality, the form of a service agreement would be shaped to some extent by the governance structure of the repository. At present, an agreement with “ReCAP” would place obligations on both ReCAP staff and ReCAP partner libraries and would thus need approval by the ReCAP Board. Ideally, the agreement would cover the collections of all ReCAP partners on similar terms, but there would be nothing to preclude an agreement that applied only to one or two partners’ materials. Once a model agreement had been implemented, the ReCAP Board might decide to authorize the ReCAP Director to execute similar agreements with other clients.

Any agreement intended to cover a partner’s collections outside of ReCAP would need separate agreements with the partner library (governing scope and policies) and with ReCAP (governing services to be provide by ReCAP itself.)

Scope:

As originally envisioned, the Cloud Library project was limited to books held at ReCAP and by NYU, and accessible through Hathi Trust (i.e., books in public domain.) At present, these
stringent requirements apply to only a small percentage of NYU’s collections. The value of an agreement could be extended if it were expanded to cover:

- All Hathi Trust books held at ReCAP and by NYU, regardless of copyright status, if the digital copy is accessible by other means (such as a Google Book Search subscription database)
- All of the ReCAP Partners’ collections also held by NYU and Hathi, regardless of whether the books are currently stored at ReCAP.
- All digitized books held at ReCAP and by NYU and available through any trusted digital repository (e.g., Portico).
- Any combination of the above.

Each of these expansions would affect the nature and terms of any service agreements.

**Terms of use:**

As noted above, the client library would want level of access and service greater than that available through standard interlibrary loan. This may be achieved through a combination of several factors:

a. The ability for the client’s patrons to discover holdings and place requests via the client’s catalog;

b. Expedited delivery;

c. Extended loan periods;

d. Availability of materials excluded from general interlibrary loan.

Ideally, the client would want a level of access similar to what it would provide through a locally-managed remote storage facility.

ReCAP partner libraries would want to ensure that extending access to client libraries would not cause a significant deterioration of service to their own patrons. As an experiment, ReCAP might be willing to extend liberal terms of use, on the assumption that ready availability of digital copies will further reduce demand for these already low-use books, both by the clients’ patrons and by the owning library. An initial agreement might include provisions for monitoring the level of use and adjusting terms if necessary.
Operational issues and responsibilities:

Defining eligible materials:

The agreement would broadly define the scope of materials covered. It might, for example, cover:

a) all books stored at ReCAP as of the date of the agreement or thereafter, except those with borrowing restrictions, if;

b) the book was also owned by the client as of the date of the agreement, and;

c) a digital copy is freely available to the client’s users.

Acting on this definition would require the parties to compile, share, and maintain data. Responsibility might be placed with either party, or contracted by mutual agreement to a third party. For a generalized agreement—one applicable to multiple clients—ReCAP might agree to provide a periodic list of eligible books, contracting with OCLC to analyze overlap with Hathi Trust for example. Client libraries might take responsibility for analyzing overlap with their own collections, or might contract that to ReCAP (and indirectly to OCLC) for an additional fee.

From the client standpoint, only those books owned as of the agreement date would be considered eligible. So, ReCAP might prefer to maintain a file of those titles, and notify clients periodically of any books that become subject to the agreement as a result of additions to ReCAP or Hathi Trust holdings.

Requesting materials:

ReCAP would be responsible for providing clients with sufficient information to place requests, but “sufficient information” could have different meanings, with different costs:

- ReCAP could supply both bibliographic information and barcode numbers, and require clients to submit requests in a standard format including the barcode number, for automated processing.

- Or, ReCAP could supply only bibliographic information and receive and process requests in a manner similar to that for interlibrary loan.

The latter method allows the client greater freedom, but would incur greater cost. (It should be noted that ReCAP itself is not in a position to supply bibliographic information directly; that information would be compiled from the ReCAP partners or possibly OCLC.)
ReCAP might also assume responsibility for notifying clients when items are in use elsewhere and therefore unavailable. Alternatively, ReCAP could provide this information only when such an item is requested. Given the expected low use, the latter may be more cost-effective.

**Delivering materials:**

ReCAP would commit to a specific turnaround for filling requests—most probably, one business day. ReCAP does not currently operate a courier service; instead, each partner is responsible for arranging to pick up and return materials. Interlibrary loans are shipped by UPS. For the Cloud Library agreement, ReCAP might offer several delivery options at different costs. Alternatively, a client might contract separately with one of the ReCAP partners for delivery.

**Terms of use:**

A Cloud Library agreement would define the terms under which requested items could be used: length of loan, renewals, right to recall items, etc. Given the expected low demand, ReCAP’s partners might agree to terms similar to those extended to their own patrons. More than one level might be defined, allowing some items to be used only on site in the client library, for example, so that the agreement could be extended to items not generally available for circulation.

ReCAP itself does not have any mechanism to control terms of use once an item leaves the facility; in effect, all items are supplied on indefinite loan. It might be left to each partner to devise its own means for enforcing limited loan periods, recalling needed items, etc. Alternatively, these activities might be added to the responsibilities of ReCAP’s interlibrary loan staff, at additional cost.
References


---

\(^1\) This estimate is based on median figures for Volumes Added (Gross) as a percentage of Total Volumes in Library as reported in the ARL Annual Statistics Tables for the five years from 2003/2004 through 2007/2008. http://www.arl.org/stats/annualsurveys/arlstats/statxls.shtml.

\(^2\) A mapping of OCLC Conspectus divisions to respective Dewey Decimal, Library of Congress and NLM call numbers is available here: http://www.oclc.org/collectionanalysis/support/conspectus.xls.