Objectives
The ReCAP partner institutions (The New York Public Library (NYPL), Columbia University, and Princeton University) have undertaken a planning project, with support from The Andrew W. Mellon Foundation, to transform the nature of the partnership from management of a shared space at the ReCAP facility to management of a shared collection. The goal of this effort is to ensure that the shared collection is managed efficiently, preserved appropriately, and available through seamless discovery and delivery to patrons of each of the three partner libraries. Another key outcome of the project will be to position ReCAP to take on a broader role within a national network of shared-print repositories.

Primary project objectives included 1) defining policies to govern, manage, and share costs of a ReCAP shared collection and 2) developing a systems architecture plan to support the desired collections management, collections development, and discovery-to-delivery functions.

The details of the planning project and the partners’ work to realize their objectives are outlined below. Most significantly, however, was the unanimous endorsement by the the ReCAP Board in September 2013 of the principles of the shared collection as outlined below and in the appendices to this report, and the approval to commence work on full implementation of these principles.

Deliverables
The planning project was organized around several activities, each with its own deliverable(s). ReCAP partners engaged several consultants to support these planning activities as identified below. All activities involved substantial work and review by the ReCAP Planning and Policy Committee (see Appendix A).

f. Workflow and technology requirements. Consultants: Sustainable Collections Services (SCS); Marshall Breeding; HTC Global, Inc.

Accomplishments
The ReCAP Planning Committee, consultants, and directors of ReCAP partner libraries successfully accomplished the planning activities identified for the grant period and produced an overall plan for implementing a ReCAP Shared Collection with Discovery/Delivery System. Descriptions for each activity are outlined below, and individual deliverable reports are available in the appendices.

Strategic planning
The ReCAP Planning Committee and Planning Consultant Lizanne Payne conducted strategic planning efforts in the early months of the project to establish shared understandings and priorities. This activity
culminated in Ms. Payne’s presentation to the Planning Committee and ReCAP Library Directors on May 18, 2012, at which the group endorsed the following shared priorities and project scope:

- Define the ReCAP Shared Collection as designated materials that are physically housed at the ReCAP facility (present and future) and covered by a ReCAP ownership/retention agreement
- Focus on discovery and access for the ReCAP Shared Collection as defined above but not for partner general collections (outside of ReCAP holdings) (not a goal of this project);
- Consider cooperative collection development (future acquisitions) of general collections material later but not as a goal of this project.

Deliverable provided: ReCAP Planning Report (Appendix B)

**Holdings analysis**

The ReCAP partner libraries engaged OCLC Research to analyze partner collections and current holdings in the ReCAP facility during 2012. This collection analysis provided a number of reports that supported subsequent decisions about the shared collections.

OCLC Research (led by Constance Malpas) provided three sets of reports including two major rounds of new analysis for the ReCAP shared collections project:

- Review of earlier data from the Cloud Library project
- “Pre-grant analysis” in March 2012, designed to be completed before the current planning project began in April 2012
- Detailed analysis using updated data in July-October 2012

OCLC Research provided a variety of comparisons between ReCAP partner collections, ReCAP holdings, Hathi Trust holdings, and WorldCat holdings. Two of the most significant findings were:

- Over 4.8 million monograph titles and over 245,000 serial titles were reported as potentially shareable ReCAP holdings
- There is significant overlap (42 percent) between current ReCAP storage inventory and HathiTrust. By effectively surfacing access to digital surrogates -- including search-only access for in copyright volumes -- ReCAP partners may be able to reduce retrievals and associated operating costs.
- Based on data supplied, title-level duplication in current ReCAP inventory is calculated at 11 percent. Rather than de-duplicating this stock, ReCAP might instead view it as a seed-collection for a shared print offering beyond the ReCAP partnership.

Deliverable provided: ReCAP Holdings Analysis Report by consultants Malpas and Payne (Appendix C)

**Policy changes**

The Planning Committee worked with Planning Consultant Payne to develop new mechanisms for governance, ownership, and permanence of collections held at ReCAP, along with related policies to manage the Shared Collection. The committee defined the following policies:

- Retention: The owning library agrees to maintain Shared Collection materials in ReCAP for the duration of the ReCAP Consortium Agreement (e.g. indefinitely).
- Ownership: Items in the ReCAP Shared Collection will remain the property of the original owning library.
- Duplication: Duplicate items may not be added to the Shared Collection, but may be added to a ReCAP partner’s Open or Restricted Collections.
- Disclosure: ReCAP libraries will report Shared Collection holdings (including retention commitments) to OCLC and will monitor and adopt, as appropriate, emerging national standards and guidelines related to shared print collections.
• Circulation: Most Shared Collection materials may circulate to authorized borrowers of any ReCAP partner, including circulation of the physical volume. Owning libraries may designate certain materials to be restricted to in-library use at the borrowing library, at the owning library’s discretion. Some materials will not circulate or will be accessible only under very limited conditions, i.e. a dark archive. (Note: This category is being provided for future use. There are no Shared Collection materials at present that are expected to be held as a dark archive.)

• Condition: Items being transferred to ReCAP for the Shared Collection must meet agreed standards for Minimum Acceptable Condition. Volumes will also be reviewed for Minimum Acceptable Condition after circulation from ReCAP

ReCAP partners wish to signal the importance of the Shared Collection by embedding its policies and commitments within the overall Consortium Agreement (originally executed in 2000). The Consortium Agreement, signed by each institution’s chief executive, defines details of the partner relationship, terms and conditions for participation, and facility policies.

Deliverables provided:
- Proposed Revisions to ReCAP Consortium Agreement (Appendix D)
- ReCAP Shared Collection Plan (Appendix E)

Cost analysis
Planning Consultant Payne prepared an analysis of costs associated with various strategic and policy options. Costs to implement and support the ReCAP Shared Collection will be affected primarily by the required technology (Discovery to Delivery system) but also by necessary operational changes and by certain policy choices. The deliverable report “ReCAP Cost Estimates” includes cost factors and estimates for most of the planned changes. The cost estimates were not developed through a specific cost analysis, but are instead based on certain cost allocation assumptions that are already in use for the ReCAP budget and cost-sharing formula.

Deliverable provided: ReCAP Cost Estimates (Appendix F)

Cost-sharing
The objective of this activity was to develop a business model for operating the ReCAP Shared Collection. To provide background, consultant Payne developed a report on business models used by other shared print programs.

ReCAP partners identified the following principles to be accomplished by a new Shared Collection cost-sharing formula:
- Encourage contributions to the Shared Collection to facilitate preservation and resource-sharing
- Discourage contributions to the Open Collections (shareable but not part of the Shared Collection) to minimize the impact on ReCAP shelving capacity, while supporting use of ReCAP space for Restricted materials when necessary (usually special collections).
- Recognize the new borrowing patterns that may emerge using the Discovery/Delivery system
- Minimize the impact of the formula change on individual ReCAP partners whose fees may increase.
- Adapt to changing usage and priorities over time through regular review of the cost-sharing formula.

Deliverables provided:
• Business models in other shared print programs (Appendix G)
• New proposed partner cost-sharing formula included in the ReCAP Shared Collection Plan (Included in the “ReCAP Planning Report,” Appendix A)

Workflow review
ReCAP partners engaged Sustainable Collections Services, Inc. (SCS) to conduct a workflow analysis of current and potential procedures to support the Shared Collection. SCS recommended a number of potential workflow changes, of which the following were deemed to be especially important by ReCAP partners:

• Revise pick cycles at the facility around requesting library: Pull all items for a given requesting library (no matter which owner or customer code), instead of pulling all items from a given owner as is done currently.
• Standardize delivery benchmarks: Establish a standard expectation for delivery across all ReCAP partners and schedule pick cycles and transport pickups to meet that benchmark.

Deliverable provided: Workflow Analysis report, including comparative workflow analysis (Appendix H)

Technology requirements
One of the most significant achievements of this planning project was the design of a system architecture to support collections management and discovery-to-delivery functions for the ReCAP Shared Collection. Consultant Marshall Breeding identified and assessed current systems in use at the ReCAP facility and at partner libraries and identified areas that would need to be modified to support a Discovery/Delivery System as envisioned for the Shared Collection.

ReCAP partners subsequently engaged HTC Global, Inc. to develop a detailed system architecture and related cost estimates for the proposed Discovery/Delivery system. As designed, the system is expected to be developed as a cloud-based middleware system searchable through each ReCAP library’s local discovery system, to provide visibility and real-time status of ReCAP Shared Collection items. The ReCAP Discovery system will also integrate with request and circulation transactions supported by ReCAP partner library systems. The ReCAP Discovery/Delivery system may be useful to other shared print programs or consortia that wish to support an integrated discovery and request environment for multiple separate systems.

Deliverables provided:
• ReCAP Project Technology Report, prepared by Consultant Marshall Breeding (Appendix I)
• ReCAP High-Level Architecture 1.0 prepared by Consultant HTC Global, Inc. (Appendix J)
• Discovery to Delivery System Cost Projections prepared by Consultant HTC Global, Inc (Appendix K)

Endorsement by the ReCAP Board
As noted above, the overall plan incorporating recommended policy changes, governance changes (revised Consortium Agreement and cost-sharing proposal), and technology recommendations was endorsed by the ReCAP Board of Governors on September 26, 2013.

Challenges
The primary challenge encountered during the project was the difficulty hiring a Technology Architect as originally planned. In the original grant proposal, the Technology Architect was envisioned as a part-time temporary staff position to be hired at the New York Public Library. After attempts to fill the position in
this way were unsuccessful, ReCAP partners modified the approach (with approval of the Foundation) to contract for system architecture services with an outside firm.

NYPL solicited bids from multiple firms and, after review by the ReCAP Planning Committee and technology staff at partner libraries, selected HTC Global to perform the planned work. HTC Global has performed similar work on the Kuali OLE system development, and thus were familiar with library and higher education applications. The difficulty in staffing the Technology Architect role resulted in a delay of several months and necessitated a no-cost extension of the grant period. The period originally planned to end on March 31, 2013 was extended to July 31, 2013.

**Project Personnel**

There were no significant changes to project personnel during the project.

**Publications**

There are no publications to date.

**Intellectual Property**

The documentation developed by HTC Global, Inc. describing the system architecture for the Discovery to Delivery System is subject to the intellectual property agreement required by the Foundation. This document is attached as the report ReCAP High-Level Architecture 1.0.

**Future Plans**

ReCAP partners are gathering additional information and soliciting bids for consulting work and system development work, with a goal of beginning implementation of the Shared Collection and Discovery/Delivery system in 2014.

**Financial Narrative**

The initial budget for the ReCAP project included expenses in two categories: project staff (salaries and fringe benefits) and consultants’ expenses. While grant expenditures remain restricted to these two categories, specific expenses changed as a result of the challenges outlined above.

As noted above, while the Library had intended to hire a Technology Architect for the project, which was unsuccessful and necessitated hiring an outside firm to perform much of the work. While the consultants for the technology architecture performed ably, The New York Public Library’s Information Technology Group staff were required to perform much of the work to coordinated the architect’s work with institutional systems. The relevant portion of their time was included on the grant. Other staff expenses for NYPL include salary and fringe benefits for Jason Bauman, NYPL’s analyst to perform collection analysis work with OCLC and ReCAP partners to identify the shared collection potential.

**Conclusion**

The project partners recognize the significance of the work performed under the auspices of this grant and are grateful to the Andrew W. Mellon Foundation for its support of this effort. The implications of the planning phase for the future of shared resources among the project partners are significant for the institutions themselves and as a model for other national print repositories.
Appendices

Appendix A. ReCAP Planning and Policy Committee Members
Appendix B. ReCAP Planning Report
Appendix C. ReCAP Holdings Analysis Report
Appendix D. Proposed Revisions to ReCAP Consortium Agreement
Appendix E. ReCAP Shared Collection Plan
Appendix F. ReCAP Cost Estimates
Appendix G. Business models in other shared print programs
Appendix H. Workflow Analysis report
Appendix I. ReCAP Project Technology Report
Appendix J. ReCAP High-Level Architecture 1.0
Appendix K. Discovery to Delivery System Cost Projections
Appendix A
ReCAP Planning and Policy Committee Members

* Ann Thornton, Andrew W. Mellon Director of the New York Public Libraries
* Jane Aboyoun, Chief Technology Officer, NYPL
* Irene David, Director, Technology Initiatives, NYPL
* Judith Johnson, Director of Enterprise Applications, NYPL
* David Magier, Associate University Librarian for Collection Development, Princeton University Library
* Marvin Bielawski, Deputy University Librarian, Princeton University Library
* Patty Gaspari Bridges, Assistant University Librarian, Collection Development, Princeton University Library
* Richard Schulz, Associate University Librarian for Technical Services, Princeton University Library
* Robert Wolven, Associate University Librarian for Bibliographic Services and Collection Development, Columbia University Libraries
* Robert Cartolano, Director, Libraries Information Technology Office, Columbia University Libraries
* Eileen Henthorne, Executive Director, ReCAP
* Lizanne Payne, Planning Consultant
This is a report required under planning activity 1d “summarizing goals, issues, and directions. The report will describe agreed priorities and directions, and will include major concerns or constraints reported by individual partners.”

**Executive Summary**

*Summary of Issues*

Several planning issues were raised during the initial site visits:

1. There is a need to clarify the meanings of “ReCAP” and “shared collection”.
2. There is an interdependency between shared collection policies and identification of materials for the shared collection.
3. There is no common resource-sharing consortium among ReCAP partners.
4. The controversy over NYPL’s Central Library Plan may affect planned transfers to ReCAP.
5. A strict NYPL deaccession policy may complicate deduping.
6. Construction at ReCAP will affect timing of Discovery to Delivery implementation.
7. The role of the Technology Architect needs to be defined and filled.

*Summary of Major Decisions*

The ReCAP Planning and Policy Committee made the following decisions on May 18, 2012.

The shared ReCAP collection has been defined as those materials housed at the ReCAP facility (or transferred there in the future) that meet certain selection criteria and are placed under a retention agreement or joint ownership agreement. Neither the criteria for selection nor the retention agreement (or joint ownership agreement) have been defined as of July 2012; these will be the subject of further discussions over the next few weeks and months.

The discovery/delivery system will be designed to provide search, display, and request functions for the shared collection at ReCAP. There is no requirement to provide a discovery system across all ReCAP partner general collections or across non-shared items at ReCAP.
Report Details

One of my first deliverables as ReCAP Planning Consultant is to “prepare a report summarizing goals, issues, and directions. The report will describe agreed priorities and directions, and will include major concerns or constraints reported by individual partners.” (Activity 1d)

The goals of the ReCAP Discovery to Delivery Project, as stated in the proposal to the Andrew W. Mellon Foundation, are as follows:

1) Reduce the cost-per-title of managing (and expanding) the shared collection;
2) Expand the breadth of material available (without Interlibrary Loan) to the users of the ReCAP partners;
3) Optimize and integrate the discovery experience in regards to the holdings of the ReCAP partnership;
4) Reduce time to delivery by streamlining request and fulfillment mechanisms; and
5) Position the ReCAP partnership to serve the needs of additional institutions.

Planning Consultant Activities (Lizanne Payne)

Between April 16 through 26, 2012, I conducted site visits at each ReCAP partner library and at the ReCAP facility, where I met with key staff to discuss project goals and local priorities or issues.

After those onsite meetings, I conducted a joint meeting of the ReCAP Planning and Policy Committee and many key staff from ReCAP partners, held at the New York Public Library on May 18, 2012. The goal of that meeting was to:

- synthesize the results of the individual meetings
- establish consensus about priorities
- identify areas of difference
- propose directions for future activities.

The report which follows incorporates the results of the original site visits and the May 18 joint meeting.

Summary of Related Consultant Activities

During this initial planning period, the other project consultants were also gathering information and performing preliminary analyses, summarized below:

1. Collection Analysis (OCLC Research)

OCLC Research performed a preliminary analysis during the pre-grant phase (February-March 2012), as stipulated by The Andrew W. Mellon Foundation, based on extracting a new set of ReCAP partner
holdings from OCLC WorldCAT, and comparing additional records identifying holdings in ReCAP and items to exclude from sharing. A Holdings Analysis Subcommittee was identified, consisting of Bob Wolven and Zach Lane (Columbia), Denise Hibay, Daphna Blatt, and Jason Baumann (NYPL), and David Magier and Patricia Gaspari-Bridges (Princeton). This group requested that OCLC perform a second round of analysis to show holdings counts by serials vs monographs, campus holdings vs ReCAP holdings, holdings in Hathi Trust (all) and in Hathi Trust public domain. This second analysis is still in process as of mid-July 2012.

2. Technology Consultant (Marshall Breeding)

Marshall Breeding conducted site visits between April 23 – 30, 2012. Key points identified during those visits were:

- There is no common ILS or discovery system among the partners
- There is no ReCAP catalog
- ReCAP uses a batch-mode GFA inventory control system with no APIs for interoperability or online status query
- Planning for new systems is underway at each partner institution, major systems are in transition

3. Workflow Consultant (Sustainable Collections Services)

SCS conducted site visits between May 8 – 14, 2012, to document current workflows for materials into and out of ReCAP and to explore what de-duplication workflows (both pre-facility and post-facility) would require.

Planning Issues

Several important issues arose during the planning discussions and with the related consultants that could affect the planning project and/or a future implementation project.

1. There is a need to clarify the meanings of “ReCAP” and “shared collection”.

   It became clear that different people meant different things by the terms “ReCAP” and “shared collection”. The proposal language stating a goal to “Optimize and integrate the discovery experience in regards to the holdings of the ReCAP partnership” was interpreted differently by various people.

   - What is ReCAP? Is it the shared facility in Princeton? The consortium of partner libraries? The consortium only in the context of the facility?
   - What is the “shared collection”? Is it only items held at the ReCAP facility? General collections at partner libraries? A “shared print” archival agreement? A resource-sharing agreement?
2. There is an interdependency between shared collection policies and selection of materials for the shared collection.

Several people mentioned that it would be important to know what the retention and access policies are before declaring certain materials as eligible for the shared collection. Conversely, the characteristics of shared materials could influence the policies defined for the shared collection. According to the proposed schedule, the content (identification of materials) for the shared collection should be completed by July 2012 while the definition of policies is part of the policy planning effort to be completed by December 2012 (although discussed in detail throughout the planning project).

3. There is no common resource-sharing consortium among ReCAP partners

Ideally it would be preferable to support lending from the shared ReCAP collection using an existing resource-sharing mechanism across the three partners. However, there is no common resource-sharing agreement in place. Columbia and Princeton belong to BorrowDirect, but NYPL does not. Columbia and NYPL belong to MARLI, but Princeton does not. It does not seem likely that these parties would join one of the existing agreements and systems.

4. The controversy over NYPL’s Central Library Plan may affect planned transfers to ReCAP.

While NYPL had made a preliminary identification of the type and quantity of materials to transfer to ReCAP in the near future, it seems likely that these planning targets may change in the wake of the public controversy over the Central Library Plan. NYPL may also face increased scrutiny of any sharing or retention agreements that may come about through the Discovery to Delivery project.

5. A strict NYPL deaccession policy may complicate deduping.

NYPL is bound by a policy that requires review and approval at a detailed level of any deaccessioning decisions. This could limit NYPL’s ability to deduplicate its collections in reliance on other ReCAP partner holdings, and could create pressures to use the NYPL copy as the shared copy of any duplicated holdings.

6. Construction at ReCAP will affect timing of Discovery to Delivery implementation.

Construction of two additional storage modules is underway at the ReCAP facility in Princeton. The timing of this construction – or, more importantly, the availability of these new modules – will affect the timing of significant aspects of the planned Discovery to Delivery systems. ReCAP partners will need an efficient mechanism to identify and prevent transferring duplicates into ReCAP by summer/fall 2013, but implementation of a new ReCAP discovery system is not likely to occur before January 2014 at the earliest, and more likely later.
Details of the timing issues are:

- Complete current planning grant: March 2013
- Complete ReCAP modules 8 & 9: June 2013
- Begin load (ReCAP ingest with significant pent-up demand): summer/fall 2013: *therefore need interim discovery tool to avoid transferring duplicates by summer/fall 2013*
- Acquire implementation funding:
  - Target for proposal: March 2013 or Sept 2013
  - Target for funding available: Sept 2013 or January 2014
- Go live with ReCAP discovery system: earliest between January – June 2014

7. The role of the Technology Architect needs to be defined and filled.

As of July 2012, the Technology Architect (included in the project proposal) had not yet been hired. Discussions among project partners indicated different understanding of the role of the Technology Architect compared to the Technology Consultant (Marshall Breeding) and lack of consensus about the level of detail that is necessary during the planning phase. The Planning and Policy Committee determined that the Technology Architect work could be postponed until after completion of the technology environmental scan, and that the project could consider engaging Marshall Breeding to perform that role as well, while continuing to recruit for the position externally.

Scope of the Shared Collection

An important question facing ReCAP planners in spring 2012 was to determine the scope of the “shared collection”. Note that a number of other planning decisions are pending and will be made during the course of the planning project.

Questions about the meaning and scope of the “shared collection” were posed and answered by the ReCAP Planning and Policy Committee during the meeting on May 18.

1. Does “Shared ReCAP Collection” = designated materials that are physically housed at the ReCAP facility (present and future) and covered by a ReCAP ownership/retention agreement? **YES.**

2. Is “discovery” required for
   - The Shared ReCAP collection as defined above? **YES.**
   - Union catalog of all partner collections? **NO.**

3. How will improved access to partner general collections be administered:
   - Through a ReCAP partners resource-sharing program? **NO.**
   - Through another existing program e.g. Borrow Direct? **MAYBE. Not a goal of this project.**
4. Will cooperative collection development (future acquisitions) of general collections material be defined through this project? **NO. Not a goal of this project.**

Questions were also raised about the importance of deduplicating holdings in various contexts:

1. Should future transfers to ReCAP be deduplicated (i.e. no new duplicates in ReCAP)? **YES.**

Items in the shared ReCAP collection should not be duplicated by future transfers to ReCAP. Further discussion will be necessary to determine the duplication policy with respect to any non-shared ReCAP items.

2. Should current ReCAP holdings be deduplicated? **NO, EXCEPT POSSIBLY FOR CONTIGUOUS JOURNALS.**

Deduplication of volumes in a high-density Harvard-model facility is a very labor-intensive and expensive process (need to retrieve, compare, reshelve, modify metadata and barcodes, dispose, and consolidate trays). It may be worth considering only for bound journals where a significant portion of a given run is shelved contiguously and therefore could release significant amounts of shelf space.

**Summary**

The shared ReCAP collection has been defined as those materials housed at the ReCAP facility (or transferred there in the future) that meet certain selection criteria and are placed under a retention agreement or joint ownership agreement. Neither the criteria for selection nor the retention agreement (or joint ownership agreement) have been defined as of July 2012; these will be the subject of further discussions over the next few weeks and months.

The discovery/delivery system will be designed to provide search, display, and request functions for the shared collection at ReCAP. There is no requirement to provide a discovery system across all ReCAP partner general collections or across non-shared items at ReCAP.
Appendix C. ReCAP Holdings Analysis Report
The ReCAP partner libraries engaged OCLC Research to analyze partner collections and current holdings in the ReCAP facility during 2012. This collection analysis provided a number of reports that supported subsequent decisions about the shared collections.

Collection Analysis Methodology

OCLC Research provided three sets of reports including two major rounds of new analysis for the ReCAP shared collections project:

- Review of earlier data from the Cloud Library project
- “Pre-grant analysis” in March 2012, designed to be completed before the current planning project began in April 2012
- Detailed analysis using updated data in July-October 2012 (see Attachment 1: ReCAP Partner Shareable Holdings Summary).

In the pre-grant analysis (March 2012), OCLC Research compared the following data sets:

- ReCAP holdings (title records) provided by each partner
- ReCAP holdings (titles, OCLC numbers) to exclude from consideration as shared collection (e.g. rare books)
- NYPL and Columbia items (title records) planned for near-term transfer to ReCAP
- Hathi Trust holdings
- WorldCat holdings

In the updated analysis (July – October 2012), OCLC Research generated new comparisons using updated ReCAP holdings (title records) provided by NYPL and Princeton. These new record sets were needed because data problems in some of the files originally provided in March 2012 adversely affected the holdings comparisons.

OCLC Research provided a variety of comparisons between ReCAP partner collections, ReCAP holdings, Hathi Trust holdings, and WorldCat holdings. Attachment 2 lists the different analyses that were performed (and files provided) in support of the ReCAP planning project. It would be very useful and not very difficult to create a separate database from the spreadsheet files provided by OCLC Research, to support further analysis of this snapshot of ReCAP holdings by subject, rights status, and other attributes.
Major collection overlap characteristics

ReCAP partners provided information on more than 4.8 million titles (OCLC records) representing items currently held in ReCAP that are potentially shareable.

All Shareable Titles in ReCAP
as of Oct 2012
N=4,851,089

- Princeton: 1,308,080 (27%)
- Columbia: 2,001,434 (41%)
- NYPL: 1,541,575 (32%)

Shareable Titles in ReCAP and Hathi
as of Oct 2012
N = 4,851,089

- ReCAP titles not in Hathi, 3,061,112 (58%)
- ReCAP in Hathi, 1,789,977 (34%)
- ReCAP in Hathi, public domain, 420,206 (8%)
Note: To get a sense of the quantity of volumes corresponding to these titles, a chart developed by Sustainable Collections Services (SCS) shows ReCAP holdings (items) for the potentially shareable customer codes:

<table>
<thead>
<tr>
<th>Customer Code</th>
<th>Customer Name</th>
<th>Total items</th>
<th>Retrieved during the last fiscal year</th>
</tr>
</thead>
<tbody>
<tr>
<td>CJ</td>
<td>Journalism Library (Columbia)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CU</td>
<td>Columbia Standard (Columbia)</td>
<td>2,576,788</td>
<td>45,196</td>
</tr>
<tr>
<td>EV</td>
<td>East Asian Vernacular (Columbia)</td>
<td>336,581</td>
<td>7,919</td>
</tr>
<tr>
<td>GC</td>
<td>Government Documents (Columbia)</td>
<td>33,422</td>
<td>264</td>
</tr>
<tr>
<td>HS</td>
<td>Health Science Library (Columbia)</td>
<td>48,447</td>
<td>700</td>
</tr>
<tr>
<td>JC</td>
<td>JSTOR Standard (Columbia)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SW</td>
<td>Social Work Library (Columbia)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PA</td>
<td>Unrestricted (Princeton)</td>
<td>1,807,683</td>
<td>19,479</td>
</tr>
<tr>
<td>QK</td>
<td>Mendel Sound &amp; Video (Princeton)</td>
<td>32,093</td>
<td>219</td>
</tr>
<tr>
<td>GP</td>
<td>Government Documents (Princeton)</td>
<td>18,183</td>
<td>273 (?)</td>
</tr>
<tr>
<td>JP</td>
<td>JSTOR Standard (Princeton)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NA</td>
<td>NYPL Standard</td>
<td>2,677,484</td>
<td>27,736</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td></td>
<td><strong>7,512,498</strong></td>
<td><strong>102,665</strong></td>
</tr>
</tbody>
</table>
Observations regarding ReCAP holdings

Upon reviewing results of the holdings analysis, Constance Malpas of OCLC Research made the following observations with respect to partner holdings in ReCAP. (These observations originally were made with respect to the March 2012 data and have been updated with data from summer/fall 2012.)

- There is significant overlap (42%) between current ReCAP storage inventory and HathiTrust. By effectively surfacing access to digital surrogates -- including search-only access for in copyright volumes -- ReCAP partners may be able to reduce retrievals and associated operating costs.

- Based on data supplied, title-level duplication in current ReCAP inventory is calculated at 11%. Rather than de-duplicating this stock, ReCAP might instead view it as a seed-collection for a shared print offering beyond the ReCAP partnership. Especially in the context of Hathi shared print plans, it seems that this additional buffer of duplication may have business value. Practically speaking, duplication in the print storage collection provides both valuable redundancy from a preservation perspective and constitutes a more reliable source of supply. This is not to say that additional duplication should be built into prospective transfers; unless the ‘business value’ of the redundancy is made real, every additional copy in ReCAP will effectively represent an opportunity cost. It’s possible that different strategies will be needed to manage retrospective redundancy in ReCAP and prospective development of the shared collections.

- A substantial portion of the near-term transfers (~20%) are already present in the ReCAP inventory. This suggests that without additional coordination in selection protocols, duplication in ReCAP inventory will continue to grow.

- An even greater part (~45%) of the near-term transfers are duplicated in HathiTrust. While relatively few of these titles are in the public domain, the aggregate library holdings per digitized title are relatively high, which may suggest that ReCAP partners can apply stricter non-duplication rules for storage transfers, maximizing the value of current investment in HathiTrust.

- The relatively high ‘average library holdings per digitized title’ in ReCAP (ranging from 131 for all digitized titles in ReCAP to 177 for digitized titles deposited by all three libraries) suggests a potentially broad market for service, should ReCAP opt to make the shared collection available to other libraries.

- Overall, the high rates of duplication with HathiTrust in extant storage inventory and near-term transfers suggests that ReCAP partner libraries will benefit from factoring the additional redundancy of HathiTrust into shared print management plans.
These totals are based on OCLC Research Library Partner reports and additional analysis prepared by OCLC Research in July - October 2012 using data on potentially shareable holdings provided by ReCAP partners.

Please note that these data were derived at different times and thus are not precisely comparable, but should be sufficient for general analysis.

<table>
<thead>
<tr>
<th>Title Records (OCLC numbers)</th>
<th>Data as of</th>
<th>Columbia</th>
<th>NYPL</th>
<th>Princeton</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Partner Collections</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partner collections Jan 2012</td>
<td>4,959,575</td>
<td>7,264,296</td>
<td>4,495,371</td>
<td>16,719,242</td>
<td></td>
</tr>
<tr>
<td>Titles in Hathi Jan 2012</td>
<td>1,669,459</td>
<td>1,903,931</td>
<td>1,554,404</td>
<td>5,127,794</td>
<td></td>
</tr>
<tr>
<td>Titles in Hathi public domain</td>
<td>246,861</td>
<td>376,918</td>
<td>316,554</td>
<td>940,333</td>
<td></td>
</tr>
<tr>
<td>Monographs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partner monographs Sep 2012</td>
<td>4,677,013</td>
<td>6,338,896</td>
<td>4,204,591</td>
<td>15,220,500</td>
<td></td>
</tr>
<tr>
<td>Monos in Hathi Jul 2012</td>
<td>1,625,868</td>
<td>1,822,665</td>
<td>1514966</td>
<td>4,963,499</td>
<td></td>
</tr>
<tr>
<td>Monos in Hathi public domain</td>
<td>254,135</td>
<td>381,381</td>
<td>332,604</td>
<td>968,120</td>
<td></td>
</tr>
<tr>
<td>Mono triplicates Sep 2012</td>
<td>1,124,643</td>
<td>1,124,643</td>
<td>1,124,643</td>
<td>3,373,929</td>
<td></td>
</tr>
<tr>
<td>Mono triplicates in Hathi Jul 2012</td>
<td>634,835</td>
<td>634,835</td>
<td>634,835</td>
<td>1,904,505</td>
<td></td>
</tr>
<tr>
<td>Mono triplicates in Hathi public domain Jul 2012</td>
<td>71,240</td>
<td>71,240</td>
<td>71,240</td>
<td>71,240</td>
<td></td>
</tr>
<tr>
<td>Serial Titles</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partner serials Sep 2012</td>
<td>146,874</td>
<td>395,955</td>
<td>170,387</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serials in Hathi Jul 2012</td>
<td>47,401</td>
<td>77,603</td>
<td>60,950</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serial triplicates Sep 2012</td>
<td>26,595</td>
<td>26,595</td>
<td>26,595</td>
<td>79,885</td>
<td></td>
</tr>
<tr>
<td>Serial triplicates in Hathi</td>
<td>18,501</td>
<td>18,501</td>
<td>18,501</td>
<td>45,503</td>
<td></td>
</tr>
<tr>
<td>ReCAP Titles (Shareable)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ReCAP titles Oct 2012</td>
<td>2,001,434</td>
<td>1,541,575</td>
<td>1,308,080</td>
<td>4,851,089</td>
<td></td>
</tr>
<tr>
<td>ReCAP in Hathi Sep 2012</td>
<td>748,375</td>
<td>512,697</td>
<td>528,905</td>
<td>1,790,976</td>
<td></td>
</tr>
<tr>
<td>ReCAP in Hathi public domain</td>
<td>147,095</td>
<td>96,132</td>
<td>176,979</td>
<td>320,206</td>
<td></td>
</tr>
<tr>
<td>Monograph titles in ReCAP</td>
<td>Oct 2012</td>
<td>1,866,851</td>
<td>1,349,041</td>
<td>1,221,654</td>
<td>4,437,546</td>
</tr>
<tr>
<td>ReCAP monos in Hathi Jul 2012</td>
<td>718,992</td>
<td>470,238</td>
<td>505,873</td>
<td>1,695,103</td>
<td></td>
</tr>
<tr>
<td>ReCAP monos in Hathi public domain Jul 2012</td>
<td>140,909</td>
<td>86,169</td>
<td>167,131</td>
<td>393,396</td>
<td></td>
</tr>
<tr>
<td>Serial Titles in ReCAP</td>
<td>62,177</td>
<td>132,432</td>
<td>51,287</td>
<td>245,896</td>
<td></td>
</tr>
<tr>
<td>ReCAP serials in Hathi Sep 2012</td>
<td>20,332</td>
<td>29,980</td>
<td>22,173</td>
<td>72,485</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Columbia</td>
<td>NYPL</td>
<td>Princeton</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>----------</td>
<td>------</td>
<td>-----------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td><strong>ReCAP Duplicate Titles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ReCAP Duplicate Titles</td>
<td>Sep 2012</td>
<td></td>
<td></td>
<td>471,593</td>
<td></td>
</tr>
<tr>
<td>ReCAP duplicates in Hathi</td>
<td>Sep 2012</td>
<td></td>
<td></td>
<td>265,483</td>
<td></td>
</tr>
<tr>
<td>ReCAP duplicates in Hathi public domain</td>
<td>Sep 2012</td>
<td></td>
<td></td>
<td>53,097</td>
<td></td>
</tr>
<tr>
<td>ReCAP mono duplicates</td>
<td>Sep 2012</td>
<td></td>
<td></td>
<td>442,422</td>
<td></td>
</tr>
<tr>
<td>ReCAP mono duplicates in Hathi</td>
<td>Sep 2012</td>
<td></td>
<td></td>
<td>251,656</td>
<td></td>
</tr>
<tr>
<td>ReCAP mono duplicates in Hathi public domain</td>
<td>Sep 2012</td>
<td></td>
<td></td>
<td>48,940</td>
<td></td>
</tr>
<tr>
<td>ReCAP serial duplicates</td>
<td>Sep 2012</td>
<td></td>
<td></td>
<td>18,958</td>
<td></td>
</tr>
<tr>
<td>ReCAP serial duplicates in Hathi</td>
<td>Sep 2012</td>
<td></td>
<td></td>
<td>11,004</td>
<td></td>
</tr>
<tr>
<td>ReCAP serial duplicates in Hathi public domain</td>
<td>Sep 2012</td>
<td></td>
<td></td>
<td>4,135</td>
<td></td>
</tr>
<tr>
<td><strong>ReCAP Triplicate Titles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ReCAP triplicates</td>
<td>Oct 2012</td>
<td>48,880</td>
<td>48,880</td>
<td>48,880</td>
<td></td>
</tr>
<tr>
<td>ReCAP triplicates in Hathi</td>
<td>Sep 2012</td>
<td>32,370</td>
<td>32,370</td>
<td>32,370</td>
<td></td>
</tr>
<tr>
<td>ReCAP triplicates/Hathi/public domain</td>
<td>Sep 2012</td>
<td>6,480</td>
<td>6,480</td>
<td>6,480</td>
<td></td>
</tr>
<tr>
<td>ReCAP mono triplicates</td>
<td>Oct 2012</td>
<td>46,575</td>
<td>46,575</td>
<td>46,575</td>
<td></td>
</tr>
<tr>
<td>ReCAP mono triplicates in Hathi</td>
<td>Sep 2012</td>
<td>30,858</td>
<td>30,858</td>
<td>30,858</td>
<td></td>
</tr>
<tr>
<td>ReCAP mono triplicates in Hathi public domain</td>
<td>Sep 2012</td>
<td>5,933</td>
<td>5,933</td>
<td>5,933</td>
<td></td>
</tr>
<tr>
<td>ReCAP serials triplicates</td>
<td>Oct 2012</td>
<td>2,254</td>
<td>2,254</td>
<td>2,254</td>
<td></td>
</tr>
<tr>
<td>ReCAP serials triplicates in Hathi</td>
<td>Sep 2012</td>
<td>1,475</td>
<td>1,475</td>
<td>1,475</td>
<td></td>
</tr>
<tr>
<td><strong>ReCAP Duplicate Monographs by owners</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Columbia/Princeton duplicate monos</td>
<td>Oct-12</td>
<td>189,321</td>
<td>n/a</td>
<td>189,321</td>
<td>189,321</td>
</tr>
<tr>
<td>ReCAP duplicates in Hathi</td>
<td>Oct-12</td>
<td>96,409</td>
<td>n/a</td>
<td>96,409</td>
<td>96,409</td>
</tr>
<tr>
<td>ReCAP duplicates in Hathi public domain</td>
<td>Oct-12</td>
<td>26,400</td>
<td>n/a</td>
<td>26,400</td>
<td>26,400</td>
</tr>
<tr>
<td>Columbia/NYPL duplicate monos</td>
<td>Oct-12</td>
<td>174,177</td>
<td>174,177</td>
<td>n/a</td>
<td>174,177</td>
</tr>
<tr>
<td>ReCAP duplicates in Hathi</td>
<td>Oct-12</td>
<td>66,868</td>
<td>66,868</td>
<td>n/a</td>
<td>66,868</td>
</tr>
<tr>
<td>ReCAP duplicates in Hathi public domain</td>
<td>Oct-12</td>
<td>2,608</td>
<td>2,608</td>
<td>n/a</td>
<td>2,608</td>
</tr>
<tr>
<td>NYPL/Princeton duplicate monos</td>
<td>Oct-12</td>
<td>n/a</td>
<td>100,172</td>
<td>100,172</td>
<td>100,172</td>
</tr>
<tr>
<td>ReCAP duplicates in Hathi</td>
<td>Oct-12</td>
<td>n/a</td>
<td>15,930</td>
<td>15,930</td>
<td>15,930</td>
</tr>
<tr>
<td>ReCAP duplicates in Hathi public domain</td>
<td>Oct-12</td>
<td>n/a</td>
<td>1,609</td>
<td>1,609</td>
<td>1,609</td>
</tr>
</tbody>
</table>
Attachment 2
ReCAP Holdings Analysis and Data Sets Provided by OCLC Research

These data sets were created through collection analysis by OCLC Research and were provided by OCLC Research. The individual data sets are posted on the ReCAP wiki or in a shared DropBox account as identified below.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>File Name</th>
<th>Data As Of Date</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monographs in Partner Collections and also in Hathi Trust public domain</td>
<td>testNYPpdBKSJul2012 testPULpdBKSJul2012 testZCUpdBKSJul2012 testZCHpdBKSJul2012 testZCLpdBKSJul2012 (these are labeled &quot;test&quot; because of the format; the data is complete)</td>
<td>July 2012</td>
<td>DropBox ReCAP data</td>
</tr>
<tr>
<td>Monographs held by all (triplicates) in Hathi public domain</td>
<td>PULZCUNYPdBksincommonJul2012)</td>
<td>July 2012</td>
<td>DropBox ReCAP data</td>
</tr>
<tr>
<td>Serials held by all (triplicates) in Hathi public domain</td>
<td>testPULZCUNYPHathiTrustSERincommonJul2012</td>
<td>July 2012</td>
<td>DropBox ReCAP data</td>
</tr>
<tr>
<td>Partner Library holdings in ReCAP and also in Hathi</td>
<td>Columbia_recap_HT_Jul2012.xlsx NYPL_recapHTJul2012.xlsx Princeton_recap20120815_HTJul2012</td>
<td>July/August 2012</td>
<td>ReCAP Project Wiki</td>
</tr>
<tr>
<td>Duplicates in ReCAP</td>
<td>2ReCAPOCNSep2012WCdetailscomplete</td>
<td>September 2012</td>
<td>ReCAP Project Wiki</td>
</tr>
<tr>
<td>Duplicates in ReCAP and Hathi</td>
<td>2copies inReCAPSep2012 and HT Jul 2012</td>
<td>July/Sept 2012</td>
<td>ReCAP Project Wiki</td>
</tr>
<tr>
<td>Triplicates in ReCAP</td>
<td>3ReCAPOCNSep2012WCdetailscomplete</td>
<td>Sep 2012</td>
<td>ReCAP Project Wiki</td>
</tr>
<tr>
<td>Triplicates in ReCAP and Hathi</td>
<td>3copies inReCAPSep2012 and HT Jul 2012</td>
<td></td>
<td>ReCAP Project Wiki</td>
</tr>
</tbody>
</table>
Appendix D. Proposed Revisions to ReCAP Consortium Agreement
ReCAP Consortium and Shared Collection Agreement


RECITALS

Whereas NYPL, Columbia, and Princeton are Institutions of the Research Collections and Preservation Consortium, Inc. (“ReCAP”), a New Jersey not-for-profit corporation (“the Corporation”) established by an Agreement dated June 26, 2000 (the “Consortium Agreement”) and Certificate of Incorporation and Bylaws dated May 26, 2000;

WHEREAS, the Institutions each have extensive collections of library materials to which substantial additions are continually being made;

Whereas, the Institutions operate through ReCAP a shared high density library collections storage facility (“the Facility”) in which they have deposited millions of volumes and other materials belonging to the Institutions’ library collections;

Whereas, the Institutions wish to establish a shared research collection for use by scholars, to preserve those materials through shared stewardship, and to promote effective use of space in the Institutions’ libraries and at the Facility;
WHEREAS, in addition to collaborating about the Facility, the Institutions wish to collaborate about the development and implementation of methods and systems for the storage, shared usage, discovery and retrieval of library materials and other information resources;

NOW, THEREFORE, in consideration of the mutual covenants contained herein, it is agreed as follows:

I. FACILITY AND SERVICES

1. Institutions may deposit library materials into the facility for individual Institution collections or for the Shared Collection. Institution collections are those that are 1) available to all ReCAP partners but for which no retention commitment has been made (“Open Collection”) and 2) available only or primarily to the owning Institution (“Restricted Collection”). The Shared Collection is defined in Section II: Shared Collection below.

2. Space in the Facility shall be allocated between individual Institution collections and Shared Collection space as approved by the Board. This allocation may be modified by the Board based on changing needs.

3. Except as otherwise provided herein, library materials stored or held in the Facility remain the property of the Institution which deposited the material, unless that Institution otherwise agrees in writing.
4. Prior to the commencement of any construction increasing the storage capacity of the Facility, or at any other time, the Institutions shall agree on the allocation of storage capacity to be created by that phase of construction as well as the schedule for processing and loading in of the materials of each member involved in such phase.

5. The Library Advisory Committee (“LAC”) shall oversee operations of and services provided by the Facility, and shall advise the Board on matters related to the Facility and its services, including the Shared Collection.

6. All Consortium facilities developed by and services provided by the Corporation and utilized in common shall be planned, operated and maintained in a manner compatible with the purposes and functions of the Consortium. Any proposed use or modification of the Facility which materially affects or potentially materially affects the quality, utilization, effectiveness, or the cost of operating the Facility or the Site, will be subject to the approval of the Board.

7. It is the intent of the Institutions that all decisions relating to the Consortium and the Facility are to be made in a collegial manner. Except as provided in Section IV: Withdrawal, Breach, Expulsion, any dispute among Institutions and/or Associate Members shall be resolved by the Board; and, should the Board be unable to resolve a dispute, the dispute shall be referred to the Institution’s Presidents for resolution.
II. SHARED COLLECTION

8. The Institutions agree to establish a program to retain, share and collaboratively manage certain library holdings (the “Shared Collection”). The LAC will recommend policies to the Board to govern services and operation of the Shared Collection.

9. Selection and designation of Shared Collection Materials

A. Materials that are covered by the terms defined herein will be known as the “Shared Collection”.

B. Materials will be identified for the Shared Collection under this Agreement based on selection criteria adopted or modified by the Board.

C. Institutions will have the right and opportunity to exclude individual items that otherwise would be designated as part of the Shared Collection under the approved selection criteria.

D. Institutions may designate certain other materials they have deposited in the Facility as “Open Collections” (available to patrons of other ReCAP Institutions) or “Restricted Collections” (not available to patrons at other ReCAP Institutions). Materials that fall into either of these categories will not be covered by a retention commitment.

E. Institutions may not transfer an item to the Facility for the Shared Collection if it would duplicate an item already held in the Shared Collection. Such items may be transferred to the Facility for inclusion in an Institution’s Open Collection or Restricted Collection.
10. Retention commitment and ownership

A. Except as otherwise provided herein, library materials designated as belonging to the Shared Collection remain the property of the Institution that deposited the material, unless that Institution otherwise agrees in writing.

B. Institutions agree to retain items designated as belonging to the Shared Collection in the Facility or in another location designated by the Board as part of the Facility for so long as this Agreement remains in effect.

C. The Board may define a different retention period for different categories of materials, or may modify the overall retention period, through a modification to this Agreement, but in no case will the retention period survive dissolution of the Consortium.

D. Institutions agree to maintain all of the designated materials in their original, artifactual form whenever possible. In the event of loss, damage or deterioration, the Institution shall use reasonable efforts to promptly obtain replacement copies of any of the retained items. If no original copy can be obtained, a hard copy facsimile may be used.

F. If an Institution no longer wishes to retain one or more individual items which it has previously committed to the Shared Collection, the Institution may petition the Board to release its commitment to the item(s) and will be subject to the terms described in Section IV Withdrawal.
G. If an Institution withdraws or is expelled from the Consortium as provided in Section IV or if this Agreement is terminated and the Consortium dissolved as provided in Section V, disposition of Institution’s Shared Collection holdings will be governed by the terms of Section IV Paragraph 18: Obligations upon Withdrawal or Expulsion.

11. Validation of Shared Collection Materials

Institutions agree to process and visually inspect items newly transferred to the Facility for contribution to the Shared Collection according to “Validation Standards” as approved by the Board. At minimum, Institutions will inspect items to ensure their serviceable condition (i.e. physically usable).

12. Disclosure of Shared Collection Materials

Institutions will note the retention status of Shared Collection materials within the RECAP discovery/delivery system, in OCLC WorldCat, and in other catalogs and/or shared print collection registries according to “Shared Collection Disclosure Standards” as approved by the Board.

13. Access to Shared Collection Materials

Institutions agree to make the materials available to authorized borrowers in accordance with a “Shared Collection Access and Circulation Policy” as approved by the Board.
III. Costs and Funding

14. **Allocation of Costs Among the Institutions.** All costs associated with developing and operating the Site, the Facility, and the Shared Collection will be allocated among the Institutions based on a formula approved by the Board.

15. **Capital and Operating Funding.** Each Institution agrees to contribute to the Corporation the amounts necessary to construct and operate the Site and the Facility in accordance with the capital and operating budgets defined by the Board and the allocation principles described herein. Each Institution shall be obligated to make its required contributions to the Corporation for both capital and operating expenses on a schedule to be set by the Board.

IV. WITHDRAWAL OF MEMBER; BREACH; EXPULSION

16. **Withdrawal.**

A. An Institution may withdraw from the Consortium and the Corporation by giving written notice to the other Institutions hereunder of its intent to do so. Any such withdrawal shall not be effective until the later of (a) the last day of the fiscal year of the Corporation subsequent to the fiscal year in which such notice is given, (b) eighteen (18) months after the giving of such notice, or (c) the first day of the month immediately following the date on which the withdrawing Institution shall have vacated that portion of the Facility occupied by it and removed therefrom its books and other property as specified in Paragraph 19: Obligations upon Withdrawal or Expulsion.

B. Within thirty (30) days following the effective date of such withdrawal, the remaining Institutions shall appoint an appraiser by notice to the withdrawing Institution. Within ten (10) days thereafter, the
withdrawing Institution shall appoint a second appraiser by notice to the remaining Institutions. These two appraisers shall, within ten (10) days, appoint a third appraiser by notice to all parties. If the two appraisers fail to do so within the ten-day period, either party may petition the New York Office of the American Arbitration Association to appoint such third appraiser. If the withdrawing Institution does not appoint its appraiser within the required ten-day period, the appraiser appointed by the remaining Institutions shall act alone. The appraisers, by a majority vote, or the single appraiser, as the case may be shall value the interests of the withdrawing Institution in the Consortium and the Corporation as at the effective date of such withdrawal. The remaining Institutions shall then have the option, jointly, in equal shares, or in any other proportion that they may otherwise agree upon, to purchase the interests of the withdrawing Institution at any time in the future, for cash in the amount of eighty percent (80%) of the appraised value of the interests of the withdrawing Institution as established in the foregoing manner. Such withdrawing Institution shall, however, be personally obligated to pay for its prorated share of operating expenses up to the effective date of its withdrawal.

C. Prior to the exercise (if ever) by the remaining Institutions of their option to acquire the interests of the withdrawing Institution (there being no outside date for such exercise), the withdrawing Institution’s ongoing prorated share of operating expenses (exclusive of usage charges), based upon the last Institutional contribution pursuant to paragraphs 14 and 15, shall be deemed a lien, in favor of the remaining Institutions, on the interests of the withdrawing Institution, provided, however, that there shall be no personal liability on the part of the withdrawing Institution for such expenses. The remaining Institutions may, at any time, foreclose such lien or exercise their option as aforesaid. The withdrawing Institution may not remove any assets of the Consortium or the Corporation without the written consent of the Board.
17. **Breach.**

If the Board determines that an Institution is in material breach of this Agreement, it may, by unanimous vote of the Board (the governors from the charged Institution not being entitled to vote), expel such Institution from the Consortium and the Corporation and remove its two governors from the Board. No Institution may be expelled without (a) first receiving (i) a detailed written notice of default, (ii) 60 days from the notice date to cure that default or, if such default is not susceptible of cure within such period, a period of time reasonably required to cure such default, provided that such charged Institution shall commence to cure in good faith such default within a 30 day period and thereafter prosecute the same diligently to completion, (iii) an opportunity to appear before the Board to respond to the notice of default and (b) the Presidents of the Institutions other than the charged Institution personally consenting in writing to the expulsion. Only after the steps described in (a) and (b) above have occurred shall expulsion be deemed to be complete.

18. **Expulsion.**

Any Institution which shall be expelled from the Consortium or the Corporation shall be deemed to have withdrawn from the Consortium and the Corporation as of the date such expulsion shall be deemed complete in accordance with paragraph 16 above. Such expelled Institution shall be deemed to be a withdrawing Institution as of that date, subject to the same procedures and consequences as if it had withdrawn as of such date, provided, however, that it shall be personally liable for its prorated share of operating expenses through the date of deemed withdrawal, and further provided that the cost of removing any library materials or other property belonging to the expelled Institution (and not removed by such expelled Institution), as well as any unpaid operating expenses or other charges attributable to
such expelled Institution, shall be deemed an additional lien on the expelled Institution’s interests in the Consortium.

19. Obligations upon withdrawal or expulsion.

If an Institution withdraws or is expelled from the Consortium, the Institution agrees to all of the following terms:

A. The withdrawing/expelled Institution will provide access to its Shared Collection materials to the other Institutions for so long as the withdrawing/expelled Institution owns the materials, whether those materials remain at the Facility or are removed to another location.

B. The withdrawing/expelled Institution will compensate the remaining Institutions for the loss of guaranteed future availability of its Shared Collection materials by paying to the Consortium an amount equal to the total amount paid by the remaining institutions to support the Shared Collection in all years since execution of this agreement.

C. The Board may waive these requirements for some or all affected materials if the Board agrees unanimously that the materials no longer need to be treated as Shared Collections.

D. A withdrawing/expelled Institution shall be obligated to remove its Institution Open Collections and Institution Restricted Collections and all other personal property no later than 180 days after withdrawal is deemed complete in accordance with paragraph 16 hereof. Any such property not removed shall be deemed abandoned and may be retained or disposed of by the Consortium, as it sees fit.

E. Such action of expulsion shall be without prejudice to the Corporation seeking recovery from the expelled Institution of any and all obligations which may be due from the expelled Institution at the time of expulsion and any further expenses incurred as a result of the expulsion (including, without limitation
the expense of storage and/or removal of the library materials or other personal property of the expelled Institution).

20. The interests in the Consortium and/or the Corporation of the Institutions are non-transferable except as explicitly provided herein.

V. TERMINATION AND DISSOLUTION

21. This Agreement may be terminated and the Consortium dissolved upon the unanimous written consent of the Institutions with the written assent of the Presidents of each. In such a case, the Institutions will take appropriate steps to cause the dissolution of the Corporation. Upon such termination and dissolution, the total assets held by the Corporation shall be liquidated, all debts of the Corporation paid, and any remaining net funds or property distributed to the Institutions, in accordance with the ratio of each Institution’s total capital contribution to the total capital contribution of all Institutions.

22. Upon dissolution of the Consortium:

A. Princeton shall have a right of first refusal, exercisable within thirty days of receiving notice from the Corporation of a bona fide third party offer, to purchase the Site and/or the Facility or that portion thereof which is offered from the Corporation; and

B. Should there be any portion of the Site purchased by the Corporation from Princeton which has not been developed and made part of the Facility at the time of such dissolution, provided that Princeton executes reciprocal easement and restrictive agreements among the Institutions acceptable to NYPL and
Columbia in their respective sole discretion reasonably exercised, sufficient to allow the two divided portions of the Site to be used and function independently, Princeton shall have the right, exercisable within 180 days of receiving notice of the intent to liquidate, to purchase such portion of the Site from the Corporation for its then fair market value as determined by an appraiser or appraisers chosen by a process agreed to by the Institutions at that time, or, in the event that the Institutions do not agree within 60 days of Princeton’s written notice of its election to purchase such portion of the Site, to be chosen by the New York Office of the American Arbitration Association on the application of any Institution.

VI. MISCELLANEOUS

23. Amendment. This Agreement may be amended only by unanimous written consent of the Institutions, and approved by the Board

24. Execution. This Agreement may be executed in any one or more counterparts, each of which shall be deemed to be an original instrument, but all of which together shall constitute but one instrument.

25. Notices. All notices, demands or requests made pursuant to, under or by virtue of this Agreement must be in writing and sent to the party to which the notice, demand or request is being made by postage prepaid, certified or registered mail, return receipt requested, by recognized courier service or by personal delivery as follows:

TO NYPL: The New York Public Library

[address]
TO Columbia: Columbia University

[address]

To Princeton: Princeton University

[address]

Any such notice, demand or request shall be deemed to have been rendered or given on the date of receipt, in the case of delivery by courier service or personal delivery, or three (3) business days after mailing.

26. Entire Agreement. This Agreement contains all of the terms agreed upon between the parties with respect to the subject matter hereof and supersedes any and all prior written or oral understandings.

27. Waiver. No waiver by any party of any failure or refusal of another party to comply with any of its obligations shall be deemed a waiver of any other or subsequent failure or refusal so to comply.

28. Successors and Assigns. This Agreement shall be binding upon and shall inure to the benefit of the parties hereto. No party shall have the right to assign any of its rights hereunder.

29. Article Headings. The headings of the various articles and paragraphs of this Agreement have been inserted only for the purposes of convenience, and are not part of this Agreement and shall not be deemed in any manner to modify, explain, qualify or restrict any of the provision of this Agreement.
30. Governing Law. This Agreement shall be governed by and in accordance with the laws of the State of New Jersey applicable to contracts made and performed wholly within that State without giving effect to the conflict-of-laws principles thereof.

IN WITNESS WHEREOF the parties, through their authorized representatives, have executed this Agreement as of the date shown above.

THE NEW YORK PUBLIC LIBRARY, ASTOR, LENOX AND TILDEN FOUNDATIONS

THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

THE TRUSTEES OF PRINCETON UNIVERSITY
ReCAP Discovery to Delivery Shared Collection Plan

Executive Summary

Goals

The ReCAP partner institutions (the New York Public Library, Columbia University, and Princeton University) have undertaken a planning project, with support from The Andrew W. Mellon Foundation, to transform the nature of the partnership from management of a shared space at the ReCAP facility to management of a shared collection. The goal of this effort is to ensure that the shared collection is managed efficiently, preserved appropriately, and available through seamless discovery and delivery to patrons and users of each of the three partner libraries. Another key outcome of the project will be to position ReCAP to take on a broader role within a national network of shared-print repositories.

Definition of the ReCAP Shared Collection

The ReCAP Shared Collection will consist of materials at the ReCAP facility that the owning libraries commit to maintain there for the agreed retention period and to manage according to other agreed Shared Collection policies. ReCAP partners may continue to hold materials at ReCAP that are not part of the Shared Collection: 1) Open Collections (available to all partners but not subject to a retention commitment or other shared policies), or 2) Restricted Collections (available only or primarily to the owning library).

Materials in the following categories will be designated as part of the ReCAP Shared Collection:

- Monographs already in ReCAP or transferred in the future unless explicitly designated for the Open or Restricted Collection by the owning library.
- Serial titles already in ReCAP or transferred in the future unless explicitly designated for the Open or Restricted Collection by the owning library.

This designation may cover as many as 4.4 million monograph titles and almost 245,000 serial titles initially. While many of these materials are included in Hathi Trust, their presence in Hathi Trust was not the determining factor in designating them for the ReCAP Shared Collection.

Shared Collection Policies

The following policies will govern the Shared Collection.

1. Retention: The owning library agrees to maintain these materials in ReCAP for the duration of the ReCAP Consortial Agreement (e.g. indefinitely).
2. Ownership: Items in the ReCAP Shared Collection will remain the property of the original owning library.
3. Duplication: No additional duplicates may be added to the Shared Collection. Duplicate items may be transferred to the ReCAP facility for the library’s Open or Restricted Collection.
4. Disclosure: ReCAP libraries will report Shared Collection holdings (including retention commitments) to OCLC and will monitor and adopt, as appropriate, emerging national standards and guidelines related to shared print collections.

5. Circulation: Most Shared Collection materials may circulate to authorized borrowers of any ReCAP partner, including circulation of the physical volume. Owning libraries may designate certain materials to be restricted to in-library use at the borrowing library, at the owning library’s discretion. Some materials will not circulate or will be accessible only under very limited conditions, i.e. a dark archive.

6. Condition: Items being transferred to ReCAP for the Shared Collection must meet agreed standards for Minimum Acceptable Condition. No condition review is required for items that are already in ReCAP. Shared Collection items that circulate will be reviewed again upon return from circulation and will be flagged for replacement if they no longer meet minimum standards.

**Discovery to Delivery System**

During the planning project, ReCAP partners and system consultants designed the architecture for a ReCAP Discovery to Delivery System. The proposed system is intended to be developed as a cloud-based middleware system and database that is integrated with each ReCAP library’s local discovery system to provide visibility, real-time status, and request functions for ReCAP Shared Collection and Open Collection items. The ReCAP Discovery system will also integrate with request and circulation transactions supported by ReCAP partner library systems.

**Governance**

The Shared Collection will be governed under the terms of the ReCAP Consortium Agreement, which will be modified to incorporate Shared Collection policies.

**Costs and Cost-Sharing Formula**

Costs to develop and implement the ReCAP Shared Collection will be incurred primarily for the discovery/delivery system and by necessary operational changes reflecting certain policy choices.

- One-time implementation costs are estimated at approximately $2.5 million (mostly for system development)
- Annual operating costs are estimated at approximately $100,000 (mostly for the system)

Partners are expected to share operating costs for the Shared Collection under a modified version of the existing ReCAP cost-sharing formula. Partners will continue to pay individually to support space and activities related to their Open Collections and Restricted Collections.

**Projected Implementation Timeline**

Implementation of the ReCAP Discovery to Delivery System and Shared Collection is expected to take two years. Current plans aim for a project to begin in July 2014 with completion expected by June 2016.
**Goals**

The ReCAP partner institutions (the New York Public Library, Columbia University, and Princeton University) have undertaken a planning project, with support from The Andrew W. Mellon Foundation, to transform the nature of the partnership from management of a shared space at the ReCAP facility to management of a shared collection. The goal of this effort is to ensure that the shared collection is managed efficiently, preserved appropriately, and available through seamless discovery and delivery to patrons and users of each of the three partner libraries. Another key outcome of the project will be to position ReCAP to take on a broader role within a national network of shared-print repositories.

**Definition of ReCAP Shared Collection**

After extensive analysis of ReCAP partner holdings and a series of discussions, ReCAP partners agreed to define the ReCAP holdings as follows. There will be three categories of materials held in the ReCAP facility:

- **Shared Collection**: Available to all partners and subject to a retention commitment and other policies agreed by the partners
- **Open Collections**: Available to all partners but not subject to a retention commitment or other Shared Collection policies
- **Restricted Collections**: Available only or primarily to the owning library

The Open Collections and Restricted Collections correspond most closely to current practice for ReCAP holdings. That is, the existing ReCAP “customer codes” identify materials that can be shared among all partners (Open) or may only circulate back to the owning library (Restricted).

The following materials will be contributed to the Shared Collection under the current plan. All quantities described below are based on a holdings analysis conducted by OCLC Research in August/September 2012 using shareable ReCAP holdings (i.e., holdings identified with customer codes indicating these items may be shared with other ReCAP partners). These totals are described as title counts because most of the analysis compared bibliographic records by OCLC number against WorldCat and Hathi Trust. Volume (item) counts of corresponding ReCAP holdings are available in some cases but are not included in this report.

1. **Monographs in ReCAP**
   a. Single monographs in ReCAP. There are over 3.9 million shareable monograph titles in ReCAP currently that were deposited by a single ReCAP partner. All of these shareable holdings will be declared part of the Shared Collection unless excluded by the owning library. All future deposits of shareable monographs by a single owner will automatically become part of the Shared Collection unless explicitly designated for the owning library’s Open or Restricted Collection. Any future deposits of monographs that match an existing Shared Collection volume will become part of the Open Collection for that depositing library.
b. Duplicate and triplicate monographs already in ReCAP. There are more than 488,000 monograph titles in ReCAP that were deposited by two or by all three ReCAP partners. For each such title, one will be declared part of the Shared Collection and the other(s) will be reassigned to Open Collections. This collection reassignment will be accomplished by an automated process using an algorithm to balance the Shared Collection monograph items roughly evenly among the ReCAP partners to the extent possible.

While many of these materials are included in Hathi Trust, their presence in Hathi Trust was not the determining factor in designating them for the ReCAP Shared Collection.

ReCAP partners will have the right and the opportunity to declare their holdings in ReCAP (individually or in the aggregate) as Open or Restricted Collections (not part of the ReCAP Shared Collection) based on the institution’s own individual criteria.

2. Serials in ReCAP

There are about 245,000 serial titles represented in shareable ReCAP holdings as of fall 2012. Of those, almost 19,000 serial titles are duplicated within ReCAP, and almost 2,000 are triplicated.

Serial titles already represented in ReCAP or transferred in the future will be designated part of the Shared Collection unless explicitly designated for the owning library’s Open or Restricted Collection.

Please note that serials will be designated for the Shared Collection at the title level at this time. That is, if serial volumes are transferred to ReCAP for a title that is not yet represented at all in ReCAP, the serial and volumes will become part of the Shared Collection (unless explicitly excluded). If serial volumes are transferred to ReCAP for a title that is already represented in the Shared Collection, those volumes will be treated as part of the library’s Open Collection (not Shared) unless/until a mechanism exists to add them to the Shared title.

ReCAP partners are committed to preserving journals and envision shared journal management as a long-term work in progress. ReCAP partners will explore procedures for comparing detailed holdings and compiling complete backfiles for possible implementation in a later phase.

3. Other Shared Collection categories

Partner libraries also identified two other categories of materials that may be added to the Shared Collection in the future:

- Library-nominated holdings: Items that did not fit into the predefined Shared Collection categories but the library is willing to contribute to the Shared Collection

- Coordinated purchases: Materials that two or more partner libraries agree to acquire for deposit to the Shared Collection.
Shared Collection Policies

1. Retention

The owning library agrees to maintain these materials in ReCAP for the duration of the ReCAP Consortium Agreement (e.g. indefinitely). There was some discussion of imposing a time-limited retention period such as 25 years or 10 years, but ReCAP partners believe that the nature of the partner library collections and holdings in ReCAP warrant a very long-term commitment. The existing ReCAP Consortium Agreement contains provisions for partner libraries to remove their holdings from ReCAP under certain well-defined conditions with specific provisions. ReCAP partners believe that a retention period tied to the existence of the ReCAP consortium signals a long-term commitment to these holdings that would be important for faculty and researchers.

2. Ownership

Items in the ReCAP Shared Collection will remain the property of the original owning library. The existing ReCAP Consortium Agreement explicitly states that all volumes deposited at the ReCAP facility remain the property of the depositing library. ReCAP partners plan to continue this policy for the Shared Collection as well, to assure local constituencies of continued availability of these materials.

3. Duplication

One of the goals of this project is to use ReCAP facility space more efficiently and reduce the need for future expansion. To further that goal, ReCAP partners have defined a policy to prohibit additional duplicates in the Shared Collection.

However, in recognition of partner libraries’ need to reclaim space in campus libraries, ReCAP partners may still deposit duplicating copies in their own Open or Restricted Collections. Under the proposed cost-sharing model, partner libraries will pay separately for Open/Restricted Collections, and it is hoped that the resulting financial penalty will deter depositing duplicates.

Because of the labor-intensive and expensive actions necessary to remove duplicate volumes at the high-density ReCAP facility, no explicit efforts will be made to remove existing duplicates at this time. Deduplication of existing ReCAP volumes may be reconsidered in a future phase.

Similarly, NYPL and Columbia have already prepared materials for transfer to ReCAP that would be costly to deduplicate. All partners may continue to transfer duplicative volumes to ReCAP before implementation of the Shared Collection and its supporting Discovery to Delivery system.

4. Disclosure

A major goal of the ReCAP Shared Collection project is “to position ReCAP to take on a broader role within a national network of shared-print repositories”. To support that goal, ReCAP libraries will report Shared Collection holdings (including retention commitments) to OCLC and will monitor and adopt, as appropriate, emerging national standards and guidelines related to shared print collections.

5. Circulation

There will be three circulation categories for Shared Collection items:
- Circulating: Most Shared Collection materials may circulate to authorized borrowers of any ReCAP partner, including circulation of the physical volume.
- In-library use: Owning libraries may designate certain materials to be restricted to in-library use at the borrowing library, at the owning library’s discretion. For pre-twentieth-century materials, each owning library may define a publication date cut-off to automatically assign holdings to the in-library use category.
- Non-circulating: Some materials will not circulate or will be accessible only under very limited conditions, i.e. a dark archive. (Note: This category is being provided for future use. There are no Shared Collection materials at present that are expected to be held as a dark archive.)

6. Condition

Items being transferred to ReCAP for the Shared Collection must meet agreed standards for Minimum Acceptable Condition. No condition reporting is required. Items not meeting those standards could be added to the library’s Open Collection.

Under this policy, volumes will also be reviewed for Minimum Acceptable Condition after circulation from ReCAP, to insure that the volume is still in acceptable condition for the Shared Collection. Volumes not meeting these standards would be considered high priorities for replacement, but would not be removed from the Shared Collection since possibly the only copy.

No explicit condition review will be performed on items that are already held at ReCAP, unless/until they circulate.

Process or workflow changes

As part of the Shared Collection planning project, ReCAP partners engaged Sustainable Collections Services, Inc. (SCS) to conduct a workflow analysis of current and potential procedures to support the Shared Collection. The following workflow changes were among those recommended in the SCS report and were deemed to be especially important by ReCAP partners.

1. Revise pick cycles around requesting library

Currently pick cycles at ReCAP are organized around individual customer codes; that is, all of the requests for items belonging to Columbia are picked at the same time.

With the Shared Collection and Discovery/Delivery system, library users are more likely to request volumes belonging to other ReCAP partners. Each transport (delivery) to a ReCAP library is more likely to contain volumes from other library customer codes. Therefore it may be preferable to revise the pick cycles to pull all items for a given requesting library (no matter which owner or customer code).
2. Standardize delivery benchmarks

Currently ReCAP supports separate delivery benchmarks for each partner library (e.g. “2 business days”, “next business day by 5:00 pm”). Once the pick cycles are separated from individual library customer codes, it would be possible to establish a standard delivery benchmark and schedule pick cycles and transport pickups to meet that benchmark. A standard delivery benchmark would be easier to manage and to convey to researchers and could facilitate acceptance.

Policies and processes for later consideration

Certain policies and processes were identified during planning that may be desirable but were postponed for later phases in order to define a more manageable scope and cost for initial implementation.

1. Compile complete backfiles of serials at ReCAP with validation for completeness

For the serial titles at ReCAP that are designated as Shared Collection titles, the following activities could be performed in order to actively compile complete backfiles:

- Identify all Shared Collection titles and volumes held at ReCAP (automated process),
- Choose the deepest backfile as the base run and identify other existing volumes at ReCAP to fill gaps (semi-automated process with staff review)
- Update Shared Collection records to record backfile holdings
- Change unneeded duplicate volumes to the library’s Open collection for future disposition (staff or automated process).

2. Validate volumes for “reportable condition” problems

In addition to the Minimum Acceptable Condition standards, other condition problems could be identified that are not significant enough to preclude items from the Shared Collection but are important enough to report (via metadata) and seek better copies. This policy is being postponed for later consideration because it would require substantial effort/cost to record condition problems for a relatively high percentage of volumes.

3. Provide retrievals and delivery 7 days/week

Improved delivery may facilitate acceptance of the Shared Collection among library users. Performing retrievals and delivery on Saturdays and Sundays would require additional shifts at ReCAP for more pick cycles and refiling (overnight and/or weekends), two additional transport pick-ups for each partner library each week, and staffing for libraries to receive and process weekend deliveries.
Discovery to Delivery System

During the planning project, ReCAP partners and system consultants have designed the architecture for a ReCAP Discovery to Delivery System. The proposed system is expected to be developed as a cloud-based middleware system searchable through each ReCAP library’s local discovery system, to provide visibility and real-time status of ReCAP Shared Collection items. The ReCAP Discovery system will also integrate with request and circulation transactions supported by ReCAP partner library systems. The ReCAP Discovery system may be useful to other shared print programs or consortia that wish to support an integrated discovery and request environment for multiple separate systems.

The proposed ReCAP middleware will support the following functions (adapted from Technology Architect report):

- **Discovery:** ReCAP middleware consolidates and normalizes ReCAP item and bib records from all three partners and provides nightly feeds to all partner discovery systems. The ReCAP search service provides the ability to perform federated search on the Shared Collection from the library discovery system.

- **Real-time Availability:** ReCAP middleware database maintains real-time status of all ReCAP Shared and Open Collection items. Item availability is provided through ReCAP middleware API.

- **Real-time Request Processing:** ReCAP middleware maintains validation rules and item status. Requests submitted through library discovery system forms are validated real-time, processed and recorded in ReCAP middleware database. Users receive confirmation or validation error messages in real-time enabling them to resubmit a valid request.

- **Real-time Status Reporting:** ReCAP middleware consolidates a complete view of item status across the ReCAP inventory system and partner library systems into the middleware database. Consolidated status can be leveraged for tracking and analytics.

- **Collection Management:** ReCAP middleware implements centralized automated collection classification algorithm (including duplicate detection). Middleware provides user interfaces for manual workflow steps such as withdrawal of preservation copies.

The proposed ReCAP middleware system would be cloud-hosted but administered by one or more ReCAP partners.

**Governance**

The Shared Collection will be governed under the terms of the ReCAP Consortium Agreement, which will be modified to incorporate Shared Collection commitments and policies.

The ReCAP Bylaws establish primary governance of the shared ReCAP facility with the Board of Governors (the Board) and also explicitly define a Library Advisory Committee (LAC) “to advise the Board...on matters relating to operation of the Facility”. The Consortium Agreement, executed by each institution’s chief executive, defines details of the partner relationship, terms and conditions for
participation, and facility policies. While some consideration was given to defining the Shared Collection through a separate agreement or MOU, ReCAP partners wish to signal the importance of the Shared Collection by embedding its policies and commitments within the Consortium Agreement itself.

Projected Costs

Costs to develop and implement the ReCAP Shared Collection will be incurred primarily for the required discovery/delivery system and by necessary operational changes reflecting certain policy choices.

Discovery to Delivery System Costs

The cost estimates below were derived from work conducted during the Discovery to Delivery planning project and are intended to establish a ballpark; actual costs may vary by +/- 20% or more. Detailed implementation costs for software and consultants will be determined by RFP during implementation planning. These estimates include new direct costs only and do not include related staff time at partner libraries and the ReCAP facility.

Estimated implementation costs

<table>
<thead>
<tr>
<th>Cost Category</th>
<th>Planning Estimate</th>
<th>Source of Planning Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery/Delivery middleware</td>
<td>$2,000,000</td>
<td>Technology Architect report</td>
</tr>
<tr>
<td>development</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interfaces to partner systems</td>
<td>$200,000</td>
<td>Various vendors and related estimates</td>
</tr>
<tr>
<td>Technology consultant (two years)</td>
<td>$150,000</td>
<td>Derived from Technology Consultant costs</td>
</tr>
<tr>
<td>Project management consultant</td>
<td>$150,000</td>
<td>Derived from Planning Consultant costs</td>
</tr>
<tr>
<td>(two years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated one-time total</td>
<td>$2,500,000</td>
<td></td>
</tr>
</tbody>
</table>

Estimated annual operating costs

<table>
<thead>
<tr>
<th>Cost Category</th>
<th>Planning Estimate</th>
<th>Source of Planning Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery/delivery middleware</td>
<td>$70,000</td>
<td>Technology Architect report (rounded)</td>
</tr>
<tr>
<td>support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interfaces to partner systems</td>
<td>$10,000</td>
<td>5% of implementation cost</td>
</tr>
<tr>
<td>support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional Princeton management</td>
<td>$20,000</td>
<td>Additional 10% of current fee</td>
</tr>
<tr>
<td>fee</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated annual total</td>
<td>$100,000</td>
<td></td>
</tr>
</tbody>
</table>
Proposed cost-sharing formula

The following planning assumptions were used to develop a revised cost-sharing formula to support the annual operating costs of the ReCAP Discovery System and Shared Collection. Certain other costs will be incurred related primarily to activities at the ReCAP partner libraries that are ordinarily absorbed by the libraries. Thus, these costs were not included in discussion of a proposed new cost-sharing formula outlined below.

Planning assumptions are:
- Use the existing ReCAP cost-sharing formula as a base for the ReCAP Shared Collection formula; i.e. retain the existing cost factors, do not revise definitions and allocations.
- Show the Discovery System operating costs explicitly. Do not include the one-time system implementation costs in this formula.

The original ReCAP cost-sharing formula uses the following structure:
- Costs are divided into two primary expense categories: “Accessions, retrievals, refiles” (ARR) and Storage
- ARR costs are based on Activity Units, i.e. weighted transactions attributable to individual partners, converted to a percent of all transactions
- Storage costs are based on Allocated Shelves, the percentage of ReCAP shelves occupied by or reserved for each partner’s holdings
- ReCAP budget line items (e.g. Labor, Utilities, Repairs) are allocated across ARR and Storage in various agreed proportions (such as Clerical Labor divided as 80% ARR and 20% Storage).
- The “calculated split of expenses” (cost-sharing formula) divides the ReCAP budget line items among the ReCAP partners according to their individual ARR and Storage percentages.

Several options were considered for modifying the current cost-sharing formula to accommodate the new Shared Collection and Discovery/Delivery system, including "no change" and "equal shares". After some discussion, the following approach is considered the most viable.

Proposed Shared Collection formula

ReCAP partners support the following principles to be accomplished by the new formula:
- Encourage contributions to the Shared Collection to facilitate preservation and resource-sharing
- Discourage contributions to the Open Collections (shareable but not part of the Shared Collection) to minimize the impact on ReCAP shelving capacity, while supporting use of ReCAP space for Restricted materials when necessary (usually special collections).
- Recognize the new borrowing patterns that may emerge using the Discovery/Delivery system
- Minimize the impact of the formula change on individual ReCAP partners whose fees may increase.
- Adapt to changing usage and priorities over time through regular review of the cost-sharing formula.
ReCAP partners proposed the following plan to modify the existing ReCAP cost-sharing formula to support these principles:

1. Divide costs for Storage into two subcategories:
   - Storage of the partner collections (Open and Restricted) using the current storage allocation, for now
   - Storage of the Shared Collection, divided proportionally by share of the Shared Collection
2. Assign a higher cost to the partner collections and a lower cost to the Shared Collection
3. Calculate and charge Retrievals and Refiles based on borrowing library rather than owning library (customer code)

Under this approach, ReCAP partners would continue to pay for their holdings in the Shared Collection (and would continue to own them), but their Shared Collection fees would be subsidized, in effect, by the higher amounts paid for private partner collections.

**Summary**

ReCAP partners plan to place most monographs and serials held at the ReCAP facility into a Shared Collection for long-term retention (as long as the consortium exists), including current and future shareable holdings. Initially, this designation could cover more than 4.4 million monograph titles (estimated 5 million volumes) and about 245,000 serial titles. Almost all shareable volumes already at ReCAP will be contributed to the Shared Collection. Most volumes in the Shared Collection may circulate to partner library patrons and will be available for ILL to other libraries.

ReCAP partners plan to develop a Discovery to Delivery middleware system to provide seamless access for partner library patrons to the Shared Collection and other materials available to them at ReCAP. The ReCAP Discovery system may also be useful to other shared print programs or consortia that wish to support an integrated discovery and request environment for multiple separate systems. Shared Collection materials will also be disclosed in OCLC WorldCat to support worldwide visibility of these shared print holdings.

Implementation of the ReCAP Shared Collection and Discovery/Delivery system is expected to take two years at an estimated cost of $2.5 million with ongoing operating costs of $100,000 annually.
ReCAP Shared Collection Plan
Related Documents

1. ReCAP Shared Collection Plan (this document) (Lizanne Payne)
2. Holdings Analysis: OCLC Holdings Analysis report (OCLC Research)
3. Workflow Analysis report (Sustainable Collections Services, Inc)
4. ReCAP Project Technology Report (Marshall Breeding)
5. ReCAP High-Level Architecture 1.0 (HTC Global, Inc.)
6. Proposed Revisions to ReCAP Consortium Agreement (Lizanne Payne)
7. Discovery to Delivery System Cost Projections (HTC Global, Inc)
8. ReCAP Shared Collection Cost Factors and Cost-Sharing Formula (Lizanne Payne)
Appendix F. ReCAP Cost Estimates
Cost Factors and Cost-Sharing Formula for the ReCAP Shared Collection
Lizanne Payne

Summary of ReCAP Shared Collection

ReCAP partners have defined policies, technology, and operational changes that are deemed critical to meet project goals: a ReCAP Shared Collection with seamless access for patrons of partner libraries. In addition, certain policy and operational options that could be implemented in later phases may have additional costs.

Phase 1 (baseline) system, services, and policies for the ReCAP Shared Collection:
- Selection: Shareable monographs and serials already at or transferred in future to the ReCAP facility will be declared part of the Shared Collection (unless otherwise excluded).
- Duplication: No new duplicates may be added to the Shared Collection, but may be added to a library’s Open Collection.
- Disclosure: ReCAP libraries will report Shared Collection holdings to OCLC.
- Condition Review: Volumes will be reviewed for minimum acceptable condition (similar to the review currently being performed) upon initial transfer to ReCAP and upon circulation.
- Access/Delivery: Shared Collection items will be available to all ReCAP partners for delivery to the requesting library. Related workflow and policy changes:
  - Revise pick cycles around requesting library
  - Standardize delivery benchmarks
- Technology: ReCAP partners will implement a Discovery to Delivery system to facilitate access to Shared Collection materials.

Phase 1: Shared Collection Costs

Costs to implement and support the ReCAP Shared Collection will be affected primarily by the required technology (Discovery to Delivery system) but also by necessary operational changes and by certain policy choices.

1. Costs of Policies and Related Operational Changes

Certain policies and operational changes will affect the costs of implementing and operating the ReCAP Shared Collection. Please see Attachment 1: ReCAP Costs Background for a detailed description of how ReCAP activity costs and storage costs were estimated or attributed for this analysis. Please note that these costs were not developed through a specific cost analysis, but rather are based on certain cost allocation assumptions that are already in use for the ReCAP budget and cost-sharing formula. The specific calculations used data from FY 13, the most recent complete fiscal year, but before making any
major decisions it would be useful to compare totals across multiple years to get an average or estimate of variation.

a. Duplication: No new duplicates

Policy: No new duplicates may be added to the Shared Collection. Duplicate items may be added to the library’s Open or Restricted Collection or may be withdrawn, at the library’s option. For serials, duplication will be identified only at the title level; that is, for Phase 1 no effort will be made to determine if a particular volume is at ReCAP prior to accession; if the title is already represented in ReCAP, subsequent volumes will be designated for the Open Collection (unless/until a mechanism is developed to compare volumes). Note: the proposed policy does not require removing existing duplicates from the ReCAP facility.

Cost factors

Technology: Efficient deflection of duplicates will require an automated process to detect duplicates before accession. A duplicate deflection function will be developed as part of the ReCAP middleware (called “accession algorithm”) and is mentioned in the ReCAP High-Level Architecture 1.0 document (HTC Global, Inc.). The duplicate detection system is already included in the system cost estimate (but is not costed separately).

Accessions processing: The duplicate detection process may be designed to change the duplicating items to the library’s Open collection automatically or may provide a report back to the library identifying duplicates and requiring further action. An automatic change to the library’s Open collection during ingest would have no effect on accession processing for those items (e.g. they would have been accessioned anyway). If the sending library chooses to receive a report of potential duplicates instead, significant staff time at the library could be required to review and reprocess volumes.

Space: If duplicates are deflected from the Shared Collection but added to the Open Collections as allowed under the new policy, space at ReCAP will be used that could have been saved. Assuming that the duplication rate for future accessions is the same as the current 11% duplication rate within the facility, the additional cost could be estimated as follows (using FY 13 amounts):

<table>
<thead>
<tr>
<th>Annual accessions (volumes)</th>
<th>Duplication rate</th>
<th>Duplicate/ triplicate accessions (volumes)</th>
<th>Attributed unit cost of accessioning additional volumes</th>
<th>Activity cost of accessioning future duplicate volumes</th>
<th>Attributed unit cost of storing additional volumes</th>
<th>Annual cost of storing future duplicate volumes</th>
</tr>
</thead>
<tbody>
<tr>
<td>434,081</td>
<td>11%</td>
<td>47,749</td>
<td>$ 0.81</td>
<td>$ 38,677</td>
<td>$ 0.11</td>
<td>$ 5,252</td>
</tr>
</tbody>
</table>
b. Disclosure: Report Shared Collection holdings to WorldCat

The proposed Disclosure Policy states that “ReCAP libraries will report Shared Collection holdings to OCLC and will monitor and adopt as appropriate the emerging national standards and guidelines related to shared print collections.” This policy is consistent with a goal of the ReCAP Shared Collection project: “to position ReCAP to take on a broader role within a national network of shared-print repositories”.

In April 2012, a working group from the shared print community proposed “Print Archives Metadata Guidelines” for reporting “print archives” or “shared print” holdings in OCLC WorldCat. The purpose of these guidelines is to promote worldwide sharing through WorldCat of information about long-term retention agreements for library holdings.

For ReCAP partners to disclose ReCAP Shared Collection materials using the current guidelines would require the following steps:

- Establish 3 new OCLC Symbols (e.g. NYPSP, ZCUSP, PULSP) to be assigned to partner library Shared Collection holdings at ReCAP
- Create and upload Local Holding Records (LHRs) to OCLC containing the basic retention information for Shared Collection holdings (these records can be batch-created)
- Remove the original OCLC symbol from the Shared Collection records in WorldCat (this can be processed in batch by OCLC)
- Process ILL requests for Shared Collection items in ILLIAD, WorldShare ILL, or Borrow Direct using a separate request queue for the SP symbol

Cost factors

Please note that some work to mark local holdings with retention information is necessary regardless of the desire to disclose this information to OCLC. It will be needed to support ReCAP partners’ own management of the Shared Collection in local library systems. Most of the costs outlined below are those that would be specific to the goal of reporting and managing Shared Collection holdings in OCLC WorldCat and related systems, separate from any local metadata.

OCLC cataloging: No charge for establishing new symbols. One-time setup costs for OCLC LHR batch loads for new symbols: ~$350 x 3= $1,050. Library systems staff time to generate shared print metadata in local catalogs (needed anyway) and to generate files for LHR loads. Optional: develop a function in the ReCAP Discovery/Delivery system to generate and maintain OCLC LHRs (not currently included in the high-level design). Library cataloging staff time to edit individual records (if needed) under the new OCLC symbol.

Resource-sharing: ILLIAD satellite license for new symbol ($1,200 annually) or Worldshare ILL lend-only subscription for new symbol ($300 annually). Workflow changes, training, and library staff and ReCAP staff time to process different ILL request queues for Shared Collection items.
c. Condition review (validation)

Items being transferred to ReCAP for the Shared Collection must meet agreed standards for Minimum Acceptable Condition (similar to requirements already in place). No explicit condition reporting is required. Items not meeting those standards could be added to the library’s Open Collection. No condition review is required for items that are already in ReCAP until they circulate.

Shared Collection items that circulate will be reviewed again upon return from circulation. Volumes not meeting minimum standards would be considered high priorities for replacement, but would not be removed from the Shared Collection since possibly the only copy.

Cost factors

Initial transfer to ReCAP: Workflow changes, training, and library staff time to process review for acceptable condition (but similar to procedures already in place). Workflow procedures will be needed to change the collection type to Open for any items that do not meet condition standards.

Review upon circulation: Workflow changes, training, and library staff and/or ReCAP staff time to review circulated items again and update metadata to flag for replacement. Consideration should be given to doing this review centrally at ReCAP prior to refiling, to simplify training and maximize efficiency. A ReCAP system function could be developed to automatically add the “replace” flag or note upon reading the barcode of the affected volume.

d. Access/Delivery

Shared Collection items will be available to all ReCAP partners for delivery to the requesting library. It will be necessary to change certain ReCAP procedures to accommodate this change.

   i. Revise pick cycles around requesting library

With the Shared Collection and Discovery/Delivery system, library users are more likely to request volumes belonging to other ReCAP partners. Each transport (delivery) to a ReCAP library is more likely to contain volumes from other library customer codes. Therefore it may be preferable to revise the pick cycles to pull all items for a given requesting library (no matter which owner or customer code).

Cost factors:

   • Technology: Implement changes to the GFA system to produce pick lists grouped by delivery location rather than owner
   • Workflow changes, training, and ReCAP staff time: Little or no change to workflows would be necessary if the pick lists were modified.
ii. Standardize delivery benchmarks

Currently ReCAP supports separate delivery benchmarks for each partner library (e.g. “2 business days”, “next business day by 5:00 pm”). If the pick cycles were separated from owning library customer codes, it would be possible to establish a standard delivery benchmark and schedule pick cycles and transport pickups to meet that benchmark.

Cost factors:
- Transport costs: Each ReCAP partner currently sets the schedule and pays for its pickups and returns to ReCAP. Most likely there would be no cost impact, but conceivably the transport cost might be increased for a given library if its schedule moved into rush hours instead of off-peak hours.

2. Costs of Technology: Discovery to Delivery System

The proposed ReCAP Discovery to Delivery System will support discovery of Shared Collection items, real-time availability tracking, real-time request processing, and collection management (including duplicate detection and management). The proposed system is designed as a cloud-based middleware system and database that is integrated with each ReCAP library’s local discovery system and with request and circulation transactions supported by ReCAP partner library systems. See the related report: ReCAP High-Level Architecture 1.0 (HTC Global, Inc.)

<table>
<thead>
<tr>
<th>Estimated implementation costs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost Category</strong></td>
</tr>
<tr>
<td>Discovery/Delivery middleware implementation</td>
</tr>
<tr>
<td>Interfaces to partner systems</td>
</tr>
<tr>
<td>Technology consultant (two years)</td>
</tr>
<tr>
<td>Project management consultant (two years)</td>
</tr>
<tr>
<td>Estimated one-time total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimated annual operating costs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost Category</strong></td>
</tr>
<tr>
<td>Discovery/delivery middleware support</td>
</tr>
<tr>
<td>Interfaces to partner systems support</td>
</tr>
<tr>
<td>Additional Princeton management fee</td>
</tr>
<tr>
<td>Estimated annual total</td>
</tr>
</tbody>
</table>
**Phase 2: Policies and processes for later consideration**

Certain policies and processes were identified during planning that may be desirable but were postponed for later phases in order to define a more manageable scope and cost for initial implementation.

1. **Compile backfiles of serials at ReCAP with validation for completeness**

The proposed Phase 1 policy states that all serial titles at ReCAP will be designated as Shared (unless explicitly excluded) but with no active efforts to develop complete backfiles. Under any future policy to compile backfiles, the process would require the following actions for each individual title or journal family (sequence of title changes):
   - Identify all Shared Collection titles and volumes held at ReCAP (automated process),
   - Choose the deepest backfile as the base run and identify other existing volumes at ReCAP to fill gaps (semi-automated process with staff review)
   - Update Shared Collection records to record backfile holdings
   - Change unneeded duplicate volumes to Open collection for future disposition (staff or automated process).

Note: this section does not include any actions or costs required to compare duplicate volumes for best physical condition. See the similar section below under “Remove duplicates from ReCAP”.

**Cost factors**

- Technology: reports to identify ReCAP serial holdings including deepest backfile; function to record completed backfile holdings in Shared Collection title record; function to change duplicate volumes to Open collection; function to display needed volumes for new accessions
- Workflow, training, staff: Update Shared Collection records to record volumes still needed (gaps); compare future serial volumes to existing Shared Collection volumes to determine whether to transfer to ReCAP

2. **Validate volumes for “reportable condition” problems**

In addition to the Minimum Acceptable Condition standards described for Phase 1, other condition problems could be identified that are not significant enough to preclude items from the Shared Collection but are important enough to report (via metadata) and seek better copies.

**Cost factors:** Staff time to review volumes against a list of reportable condition criteria (to be defined) and update records to report the problems found. If a “reportable condition” policy were instituted, this review could occur on initial transfer to ReCAP and optionally again after circulation of the volume.
The Western Regional Storage Trust (WEST) program uses an estimate of $4.00 per volume as the attributed unit cost for volume-level accessions with validation. This cost estimate covers physical handling and metadata updates and was derived from a cost study conducted at the University of California Shared Print Program in 2008. For ReCAP cost estimating, it would be reasonable to attribute $2.00 of that cost to the condition review portion, separate from the accessions cost.

Using that attributed unit cost, the estimated annual cost to review and report “reportable conditions” would be:

<table>
<thead>
<tr>
<th>Annual accessions (volumes)</th>
<th>Attributed unit cost of condition review</th>
<th>Estimated annual cost of reportable condition review</th>
</tr>
</thead>
<tbody>
<tr>
<td>434,081</td>
<td>$ 2.00</td>
<td>$ 868,162</td>
</tr>
</tbody>
</table>

3. Provide retrievals and delivery 7 days/week

Delivery every day may facilitate acceptance of the Shared Collection among library users.

Cost factors:
- Performing retrievals and delivery on Saturdays and Sundays would require additional shifts at ReCAP for more pick cycles and refileing (overnight and/or weekends),
- Two additional transport pick-ups for each partner library each week,
- Staffing for libraries to receive and process weekend deliveries.

4. Remove existing duplicates from ReCAP

Removing existing duplicate volumes from ReCAP would require the following actions:
- Identify duplicate (or triplicate) volumes already shelved at ReCAP
- Optional: Compare volumes for best physical condition
- Dispose of removed volumes and update records

Cost factors:
- Technology: During the initial process to create the Shared Collection, the system would designate one copy for the Shared Collection and the other(s) for the Open Collection (based on an algorithm). This process could produce a file/report for use in any future deduplicating.
- Workflow, training, staff time without condition comparison: retrieve the corresponding Open Collection volume(s), discard, update records.
- Workflow, training, staff time with condition comparison: retrieve all volumes, compare, refile the chosen volume, discard the not-chosen volumes, update records.
- Space: Holding area for volumes pending discard. With condition review, twice as much holding space is needed to stage volumes while under review, plus work space to compare volumes.
- Discard costs: Costs to transport, ship, or recycle discarded volumes.
Some costs of removing existing duplicates can be estimated using current attributed unit costs. However, please note that it is likely that unit costs would go down for a high volume of concentrated activity, i.e. in a production environment. A more detailed analysis should be performed before making a major decision about removing existing duplicates.

The following assumptions were used to calculate these estimates:

- Duplicate removal without condition comparison is composed of
  - 1 Retrieval per removed volume @ $3.38
  - 1 Withdrawal per removed volume (to update the record) = 1 Accession @ $.81
- Duplicate removal with condition comparison is composed of these additional cost factors:
  - 1 additional Retrieval to include the Shared Collection volume @ $3.38
  - 1 Condition Review per reviewed volume @ $2.00
  - 1 Refile per retained Shared Collection volume @ $3.38

Using these assumptions, the estimated one-time cost to remove existing duplicates in ReCAP (11% of current space) would be:

<table>
<thead>
<tr>
<th>Category</th>
<th>Titles</th>
<th>Est Total Volumes</th>
<th>Vols to Remove</th>
<th>Retrieve dup vols</th>
<th>Withdrawal dup vols</th>
<th>Total before Condition Review</th>
<th>Retrieval Shared vol</th>
<th>Review all vols</th>
<th>Refile chosen vol</th>
<th>Total w/ Condition Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duplicates</td>
<td>471,593</td>
<td>943,186</td>
<td>471,593</td>
<td>$ 1,593,984</td>
<td>$ 381,990</td>
<td>$ 1,975,975</td>
<td>$ 1,593,984</td>
<td>$ 1,886,372</td>
<td>$ 1,593,984</td>
<td>$ 7,050,315</td>
</tr>
<tr>
<td>Triplicates</td>
<td>48,880</td>
<td>146,640</td>
<td>97,760</td>
<td>$ 330,429</td>
<td>$ 79,186</td>
<td>$ 409,614</td>
<td>$ 165,214</td>
<td>$ 293,280</td>
<td>$ 330,429</td>
<td>$ 1,198,538</td>
</tr>
<tr>
<td>Total duplication</td>
<td>520,473</td>
<td>1,089,826</td>
<td>$ 1,924,417</td>
<td>$ 461,177</td>
<td>$ 2,385,589</td>
<td>$ 1,759,202</td>
<td>$ 2,179,654</td>
<td>$ 1,924,417</td>
<td>$ 8,248,862</td>
<td></td>
</tr>
</tbody>
</table>

Please note: this estimate covers only the activities and transaction costs to remove duplicates. It does not include any estimate to recover the space freed by those duplicates. There would be significant additional costs to consolidate the remaining trays and shelves to recover the space. Thus the estimate of over $2 million (without condition comparison) or over $8 million (with condition comparison) is BEFORE space recovery which requires costs that are almost impossible to estimate.
Proposed cost-sharing formula

The following planning assumptions were used to develop a revised cost-sharing formula to support the annual operating costs of the ReCAP Discovery System and Shared Collection.

- Use the existing ReCAP cost-sharing formula as a base for the ReCAP Shared Collection formula; i.e. retain the existing cost factors, do not revise definitions and allocations.
- Show the Discovery System operating costs explicitly. Do not include the one-time system implementation costs in this formula.

The original ReCAP cost-sharing formula uses the following structure:
- Costs are divided into two primary expense categories: “Accessions, retrievals, refiles” (ARR) and Storage
- ARR costs are based on Activity Units, actual transactions attributable to individual partners, converted to a percent of all transactions
- Storage costs are based on Allocated Shelves, the percentage of ReCAP shelves occupied by or reserved for each partner’s holdings
- ReCAP budget line items (e.g. Labor, Utilities, Repairs) are allocated across ARR and Storage in various agreed proportions (such as Clerical Labor divided as 80% ARR and 20% Storage).
- The “calculated split of expenses” (cost-sharing formula) divides the ReCAP budget line items among the ReCAP partners according to their individual ARR and Storage percentages.

Several options were considered for modifying the current cost-sharing formula to accommodate the new Shared Collection and Discovery/Delivery system, including "no change" and "equal shares". After some discussion, the following approach is considered the most viable.

Proposed Shared Collection formula

ReCAP partners support the following principles to be accomplished by the new formula:
- Encourage contributions to the Shared Collection to facilitate preservation and resource-sharing
- Discourage contributions to the Open Collections (shareable but not part of the Shared Collection) to minimize the impact on ReCAP shelving capacity, while supporting use of ReCAP space for Restricted materials (usually special collections).
- Recognize the new borrowing patterns that may emerge using the Discovery/Delivery system
- Minimize the impact of the formula change on individual ReCAP partners whose fees may increase.
- Adapt to changing usage and priorities over time through regular review of the cost-sharing formula.

ReCAP partners proposed the following plan to modify the existing ReCAP cost-sharing formula to support these principles:

1) Divide costs for Storage into two subcategories:
• Storage of the partner collections (Open and Restricted) using the current storage allocation, for now
• Storage of the Shared Collection, divided proportionally by share of the Shared Collection
2. Assign a higher cost to the partner collections and a lower cost to the Shared Collection
3. Calculate and charge Retrievals and Refiles based on borrowing library rather than owning library (customer code)

Under this approach, ReCAP partners would continue to pay for their holdings in the Shared Collection (and would continue to own them), but their Shared Collection fees would be subsidized, in effect, by the higher amounts paid for private partner collections. See Attachment 2: Possible New ReCAP Cost-Sharing Formula based on this formula.
The following cost factors and assumptions were used as background for estimating costs of certain ReCAP policy options or service options. Please note that these costs were not developed through a specific cost analysis, but rather are based on certain cost allocation assumptions that are already in use for the ReCAP budget and cost-sharing formula (described below and details attached).

Cost assumptions

Under the approach used for the ReCAP budget and cost-sharing formula, ReCAP facility costs are divided into costs for Activities and for Storage. Budget line item expenditures such as Labor, Utilities, etc. are attributed across Storage and Activities (ARR) by an agreed formula. Total ARR and Storage budgets/expenditures can then be calculated by aggregating these attributed costs.

Based on ReCAP FY 13 budget

<table>
<thead>
<tr>
<th>Activity Type</th>
<th>FY 13 Activity Units</th>
<th>Activity Unit %</th>
<th>Share of ARR cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessions</td>
<td>518,264</td>
<td>27.4%</td>
<td>$349,935</td>
</tr>
<tr>
<td>Retrievals</td>
<td>638,695</td>
<td>33.8%</td>
<td>$431,251</td>
</tr>
<tr>
<td>Refiles</td>
<td>657,200</td>
<td>34.8%</td>
<td>$443,746</td>
</tr>
<tr>
<td>Interlibrary Loans</td>
<td>76,140</td>
<td>4.0%</td>
<td>$51,410</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,890,299</strong></td>
<td><strong>100%</strong></td>
<td><strong>$1,276,343</strong></td>
</tr>
</tbody>
</table>

Further calculations can be performed to estimate attributed expenditures for individual activity units or storage units. See the attachment for detailed calculations that are summarized below.

Activities

As part of the current ReCAP budgeting process, transactions for Accessions, Retrievals, Refiles, and ILL (ARR) are weighted to normalize the activities so they can be analyzed equally. That is, projected transaction quantities are converted to Activity Units such that one “Accessions” Activity Unit is worth the same as one “Retrieval” Activity Unit.

Expenditures for individual activity types (e.g. accessions, retrievals) can then be calculated by dividing ARR expenditures proportionally for each activity type.
<table>
<thead>
<tr>
<th>Activity Type</th>
<th>FY 13 Share of ARR Cost</th>
<th>FY 13 Actual Transactions</th>
<th>Attributed Unit Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessions</td>
<td>$349,935</td>
<td>434,081</td>
<td>$0.81</td>
</tr>
<tr>
<td>Retrievals</td>
<td>$431,251</td>
<td>127,739</td>
<td>$3.38</td>
</tr>
<tr>
<td>Refiles</td>
<td>$443,746</td>
<td>131,440</td>
<td>$3.38</td>
</tr>
<tr>
<td>Interlibrary Loans</td>
<td>$51,410</td>
<td>15,228</td>
<td>$3.38</td>
</tr>
</tbody>
</table>

Reminder note: These activity type costs and attributed unit costs were not based on a detailed cost analysis of the actual work required to process these activities. Rather, these cost estimates are based on the cost allocation formula already agreed by the ReCAP partners as a method for developing and sharing the ReCAP facility costs.

Storage

Storage costs can be described in terms of two different units: shelves and items.

<table>
<thead>
<tr>
<th>Storage Unit</th>
<th>Quantity</th>
<th>FY 13 Share of Storage Cost</th>
<th>Attributed Unit Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items</td>
<td>10,230,226</td>
<td>$1,146,146</td>
<td>$0.11 per item</td>
</tr>
<tr>
<td>Shelves allocated</td>
<td>74,421</td>
<td>$1,146,146</td>
<td>$15.40 per shelf</td>
</tr>
</tbody>
</table>

Item counts are a more difficult measure because of the widely varying sizes of stored items (e.g. a volume vs an archive box). The unit cost per item is average across all types.

The attributed costs for Activity Units and for Storage are used in various calculations in the following sections to estimate the cost impact of various policy or service options. Please note that these cost estimates and impacts are very rough measures devised through general cost attribution as described here.
Attachment 2
Illustration of Possible New ReCAP Cost-Sharing Formula

The Research Collections and Preservation Consortium, Inc.
Illustration of Possible New Cost-Sharing Formula using FY 13 Budget

Under this option, costs attributed to Storage would be divided into two categories: 1) Storage of the Partner Collections (Open and Restricted) divided according to the current storage allocation, and 2) Storage of the Shared Collection divided proportionally by share of the Shared Collection. In addition, this formula assigns a higher portion of the costs to the partner collections and a lower cost to the Shared Collection.

Preliminary Illustration

<table>
<thead>
<tr>
<th>Activity</th>
<th>Shared Coll Storage (3)</th>
<th>Partner Coll Storage (3)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor</td>
<td>25%</td>
<td>79%</td>
<td></td>
</tr>
<tr>
<td>Administrative</td>
<td>50%</td>
<td>13%</td>
<td></td>
</tr>
<tr>
<td>Clerical</td>
<td>80%</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>Cleaning</td>
<td>75%</td>
<td>6%</td>
<td></td>
</tr>
<tr>
<td>Repairs &amp; Maintenance</td>
<td>50%</td>
<td>13%</td>
<td></td>
</tr>
<tr>
<td>Utilities</td>
<td>20%</td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>Roads, Grounds, &amp; Security</td>
<td>75%</td>
<td>6%</td>
<td></td>
</tr>
<tr>
<td>Administrative Insurance</td>
<td>20%</td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>All Other</td>
<td>80%</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>Major Maintenance</td>
<td>25%</td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>Discovery System (2)</td>
<td>75%</td>
<td>25%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1,352,200</td>
<td>311,700</td>
<td>2,524,000</td>
</tr>
</tbody>
</table>

(1) Accessions, Retrievals, Refiles. Retrievals/refiles based on requester, not owner. Original owner estimates used in this illustration.
(2) Discovery System expenses attributed 75% to ARR and 25% to Shared Collection Storage.
(3) Storage divided 25% to Shared Collection Storage and 75% to Partner Storage (to cover most expenses via Partner Storage)

Activity Units FY 13 Projected

<table>
<thead>
<tr>
<th>Activity</th>
<th>Princeton</th>
<th>Columbia *</th>
<th>NYPL **</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessions</td>
<td>440,000</td>
<td>208,000</td>
<td>350,000</td>
<td>998,000</td>
</tr>
<tr>
<td>Retrievals</td>
<td>120,000</td>
<td>371,000</td>
<td>186,000</td>
<td>677,000</td>
</tr>
<tr>
<td>Refiles</td>
<td>135,000</td>
<td>385,000</td>
<td>205,000</td>
<td>725,000</td>
</tr>
<tr>
<td>Interlibrary Loans</td>
<td>32,500</td>
<td>46,000</td>
<td>12,000</td>
<td>90,500</td>
</tr>
<tr>
<td>Total</td>
<td>727,500</td>
<td>1,010,000</td>
<td>753,000</td>
<td>2,490,500</td>
</tr>
</tbody>
</table>

Percent

<table>
<thead>
<tr>
<th>Princeton</th>
<th>Columbia *</th>
<th>NYPL **</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>29.2%</td>
<td>40.6%</td>
<td>30.2%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Allocated Shelves

<table>
<thead>
<tr>
<th>Activity</th>
<th>Princeton</th>
<th>Columbia</th>
<th>NYPL</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shared Coll Storage Share</td>
<td>27%</td>
<td>41%</td>
<td>32%</td>
<td>100%</td>
</tr>
<tr>
<td>Partner Storage Share</td>
<td>21%</td>
<td>42%</td>
<td>37%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Calculated Split of Expenses

<table>
<thead>
<tr>
<th>Activity</th>
<th>Princeton</th>
<th>Columbia</th>
<th>NYPL</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative</td>
<td>57,916</td>
<td>92,181</td>
<td>73,903</td>
<td>$ 224,000</td>
</tr>
<tr>
<td>Clerical</td>
<td>217,655</td>
<td>318,556</td>
<td>244,749</td>
<td>$ 781,000</td>
</tr>
<tr>
<td>Cleaning</td>
<td>5,507</td>
<td>8,171</td>
<td>6,323</td>
<td>$ 20,000</td>
</tr>
<tr>
<td>Repairs &amp; Maintenance</td>
<td>41,110</td>
<td>65,432</td>
<td>52,458</td>
<td>$ 159,000</td>
</tr>
<tr>
<td>Utilities</td>
<td>133,089</td>
<td>231,630</td>
<td>193,330</td>
<td>$ 558,000</td>
</tr>
<tr>
<td>Roads, Grounds, &amp; Security</td>
<td>32,489</td>
<td>48,207</td>
<td>37,304</td>
<td>$ 118,000</td>
</tr>
<tr>
<td>Administrative Insurance</td>
<td>17,882</td>
<td>31,133</td>
<td>25,985</td>
<td>$ 75,000</td>
</tr>
<tr>
<td>All Other</td>
<td>24,246</td>
<td>35,490</td>
<td>27,264</td>
<td>$ 87,000</td>
</tr>
<tr>
<td>PU Management Fee</td>
<td>53,584</td>
<td>83,866</td>
<td>65,531</td>
<td>$ 202,000</td>
</tr>
<tr>
<td>Major Maintenance</td>
<td>47,684</td>
<td>83,022</td>
<td>69,294</td>
<td>$ 200,000</td>
</tr>
<tr>
<td>Discovery System</td>
<td>28,658</td>
<td>40,666</td>
<td>30,676</td>
<td>$ 100,000</td>
</tr>
<tr>
<td>Total</td>
<td>659,771</td>
<td>1,037,412</td>
<td>826,817</td>
<td>2,524,000</td>
</tr>
</tbody>
</table>

Resulting percentage

<table>
<thead>
<tr>
<th>Princeton</th>
<th>Columbia</th>
<th>NYPL</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>26%</td>
<td>41%</td>
<td>33%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Comparison to current formula

| Current formula totals     | 613,523   | 968,991   | 809,975  | $ 2,422,489 |
| New formula totals         | 659,771   | 1,037,412 | 826,817  | 2,524,000   |
| per cent increase          | 7%        | 4%        | 2%       | 4%          |
| overall                   |           |           |          |             |
Appendix G. Business models in other shared print programs
This report is required under planning activity 3.

3. Determine options for policy changes, including new mechanisms for the governance, ownership, and permanence of collections held at ReCAP and if necessary of the ReCAP relationship

   a. Outline and compare governance mechanisms and policies at other library consortia with shared print collections.

Overview

A large number of library consortia or other library groups have entered into shared collection agreements over the past several years. For purposes of this document, a “shared collection agreement” refers to a formal agreement among multiple libraries to share responsibility for, management of, and access to a defined set of print materials. These agreements are variously described as “shared collections”, “shared print collections”, or “print archives”. “Shared collection(s)” will be used in this context because that is the terminology used in the ReCAP Discovery to Delivery Proposal.

Libraries and consortia developing a shared collections program typically define an agreement that covers three aspects of the collaboration:

1. Operating policies (how the shared collection functions, including most or all of the following):
   a. Selection of materials to preserve
   b. Retention commitment (period of time)
   c. Archiving location(s)
   d. Ownership
   e. Validation standards (review for completeness and condition)
   f. Holdings disclosure standards
   g. Access/delivery policies.

2. Business model (how – or if – costs are covered)

3. Governance (how decisions are made)
Agreement on all of these terms and conditions is usually embodied in a formal document, ranging from a posted statement of principles, to a Memorandum of Understanding (MOU), to a signed legal agreement.

The following section outlines the most common approaches taken by existing shared collection programs, followed by details of selected programs.

Common Attributes of Current Shared Collection Programs

Selection

“Selection” refers to the process or criteria used to identify materials that will belong to the shared collection.

The majority of current shared collections focus on journals, most likely because of these characteristics:

- Many are available in digital form, sometimes with digital preservation as well (Portico or CLOCKSS), such that use of the print version has declined significantly
- Large amounts of shelf space can be reclaimed with a decision about a single title

The most common method for selecting or identifying journals for the shared collection is by publisher or digital aggregator (e.g. “JSTOR journals” or “Elsevier journals”). Other methods include:

- Individual library-nominated titles (ASERL journal program)
- Shared storage copy becomes shared collection copy (Florida, WRLC)
- Custom collection analysis across members (WEST, COPPUL)

A few programs identify monographs in the form of “last copy” policies. If a given monograph happens to be the last copy of that work held within that group of libraries, the program defines a process for retaining it on behalf of the group. For example, the University of Illinois accepts last-copy monographs on behalf of the CARLI consortium.

A few groups are beginning to consider shared monograph collections in a more proactive way, by attempting to identify large classes of monographs to be held in a shared collection. The HathiTrust is considering program for its members to retain copies of volumes available in digital form through Hathi. The Midwest Collaborative for Library Services (MCLS) is in the early stages of implementing a shared monograph collection based on a data-driven analysis of member holdings.

Retention commitment

The retention commitment is the single most important factor in a shared collection agreement. The retention commitment is what distinguishes a shared collection from a resource-sharing agreement or from a shared storage facility. With a retention commitment, participating libraries agree to keep the designated materials for a specified (or unspecified) period of time for use by other program
participants under defined access policies. The retention commitment allows other participating libraries to consider deselecting their own copies by relying on the retained copies held by others.

Some of the earliest shared collection agreements did not specify an explicit time period, implying that the retention agreement extended indefinitely or in perpetuity. CRL’s JSTOR Archive is an example. More recent agreements tend to establish a finite time period such as 25 years or even 10 years, at which time the agreement will be reviewed and either extended or canceled.

There is a natural tension between the goals of the retaining libraries and the potentially-borrowing libraries. The borrowing libraries may prefer a longer retention period to cover their potential borrowing needs for the longest time. The retaining libraries may prefer a shorter retention period to minimize the time when their local collection management decisions are constrained. An argument often given for a near-term defined retention period (e.g. 10 or 25 years) is that digital technology may have changed so much in that time that the need for a shared print collection should be reassessed.

Archiving Location(s)

Shared collections may be established in a centralized or distributed pattern, and may involve specialized library storage facilities or traditional campus library locations. For instance, the Orbis Cascade Alliance Distributed Print Repository (DPR) program is (of course) distributed among member library locations, as is the ASERL Collaborative Journal Retention Program. The CIC Shared Print Repository currently is centralized at the Indiana University high-density library storage facility; similarly, the Five Colleges consortium in Massachusetts maintains a centralized shared collection at the Five Colleges facility.

Ownership

All shared collection programs grapple with the issue of ownership: which entity will own the physical volumes designated in the shared collection. There are three potential owners that could be identified:

- The original owning library
- The retaining library (if different)
- The consortium or program (if a separate entity).

Transferring ownership of library materials tends to be complicated. Most universities or research institutions classify library collections as assets and are likely to require explicit procedures for disposing of such assets. These procedures are often especially onerous for publicly-funded institutions.

As a result, most shared collection programs define ownership of the shared materials as remaining with the original owner. In many cases, especially for distributed collections, the original owner is also the retaining library. In some cases (e.g. WEST, Florida FLARE), if the original owner donates volumes to a retaining library, the original owner is asked to gift those volumes to the retaining library.
In a very small number of cases, the shared collection becomes the property of the consortium itself. This is true for the CRL JSTOR Archive and the Five Colleges (Mass.) Library Depository.

Validation standards

In some cases, shared collection programs are establishing validation standards to define the level of review for completeness and condition that will be performed on these materials. Validation standards support preservation goals by setting minimum standards for accepting physical items into the shared collection and standards for reporting gaps and condition problems through metadata.

At present only a small number of shared collection programs such as WEST and CIC SPR have defined explicit validation standards.

Holdings disclosure standards

Shared collection programs typically define how the shared materials will be reflected in catalog records, and which catalogs will contain those records (e.g. local OPAC, consortial OPAC or other system, OCLC WorldCAT). Especially in earlier days, this often involved detailed locally-specific guidelines for what information to record in the local ILS or consortial catalog with no attention paid to disseminating shared collection information more widely.

However, in April 2012, the OCLC Print Archives Metadata Guidelines were published by an ad hoc group that had worked to define a common metadata format for wide disclosure of print retention commitments through OCLC WorldCAT. These guidelines define a community standard for using MARC Holdings Records with 583 Action Notes to describe the retention commitment and the results of any validation that was performed. A blog post by OCLC Research provides a good description and a link to the guidelines document (http://hangingtogether.org/?p=1734)

Also, the Center for Research Libraries (CRL) is developing the Print Archives Preservation Registry (PAPR) knowledgebase, which also uses the Print Archives Metadata Guidelines as the standard for input data. PAPR will provide an online searchable registry of archived print holdings and descriptive information about shared collection programs. PAPR is expected to be released for public use in early August 2012.

Access/delivery policies

Some of the most important policies to be defined by shared collection programs are those that govern access and delivery. Shared collections need to define:

- Who will have access to the shared materials
- How will requests be made and fulfilled
- What forms of delivery will be offered
• What restrictions on use will be applied

A small number of shared collections are defining “dark archives” that are intended to preserve the materials indefinitely and make them available only in extreme circumstances. Examples include the dark archive copy of PALCI holdings at Penn State, the University of California and Harvard dark archives for JSTOR titles, and a planned JSTOR dark archive at the University of Minnesota.

Most shared collections, however, have defined “light archives” where the materials regularly are made available to partner libraries on request. Non-retaining partner libraries are better able to reclaim local shelf space if they can be assured that shared materials will be provided to them if needed; “dark archives” do not provide the same assurance of ready access.

One of the questions to be considered is whether requests will be fulfilled on a priority basis for members. The answer sometimes depends on the presence of an existing shared catalog with request feature. Programs among partners that do not share a catalog/request system may need to determine whether traditional ILL request systems will be used, or whether a new mechanism or borrowing agreement will be established. WEST member libraries decided not to provide prioritized or expedited request/delivery to WEST members because the libraries are already supporting several different resource-sharing agreements and they felt that the expected usage of WEST materials would be very low and not worth the effort of supporting a special borrowing agreement.

Shared collection programs often place special use restrictions on these materials, however. Because other libraries are depending on long-term access to the shared materials and may have deselected their own copies, shared collection programs often require “in-library use” of the physical item to minimize the chance of loss or damage.

Business Model

Each shared collection program must decide how to cover the costs of establishing, maintaining, and using the shared collection. Related decisions must be made about whether to collect funds from members to support these costs.

Participants incur costs in the following categories:

• Selection and collection analysis (staff time, systems and/or database)
• Space (indirect costs of facility financing, utilities, maintenance)
• Ingest (handling, validating, shelving volumes)
• Metadata update (changing bibliographic and holdings records per disclosure policies)
• Shipping volumes to retaining libraries
• Deselecting local volumes
• Project management and administration
The vast majority of current shared collection programs – especially those that are distributed collections -- operate on a “share/share alike” basis. In these cases, each participating library absorbs its own costs and no money changes hands. The underlying assumption is that all libraries are investing an equitable amount to the collaborative effort.

A small number of programs have developed business models in which members pay fees to support the shared collection. Both WEST and the CIC Shared Print Repository identified a set of costs that would be shared among all members as a way to support the collective goals. For WEST, shared costs include:

- Payments to Archive Builders to support ingesting and validating materials
- Use of CRL’s PAPR system for collection analysis
- Project management and administration

The business model for CIC SPR includes support for ingesting materials and also covers payments to Indiana University for use of its storage facility space.

**Governance**

Shared collection programs need to define mechanisms for determining who is a member and how decisions are made. Some form of governance supports sustainability over time when the original collaborators may no longer be involved.

Almost all shared collection programs were established within existing library consortia and rely on the existing consortium structures to provide the basic framework for shared collections governance. This is true for the CIC Shared Print Repository, ASERL Collaborative Journal Retention Program, and many others. The WEST program is unusual for being a broad-scale collaborative effort that was defined outside of existing consortial agreements. In effect WEST established a new consortium specifically to administer the shared collection.

**Agreement**

Even if the shared collection operates within an existing consortium, it is necessary to define the specific terms and conditions of the shared collection through some kind of agreement. The agreement usually identifies some or all of the following terms and conditions:

- Membership: Who are the members or participants and (optional) criteria for adding future participants
- Governance: Organizing body(ies) that will make decisions on behalf of the group
- Basic operating policies: Retention period, ownership, disclosure, access and delivery. Sometimes the main agreement refers to external documents that may contain these details.
- Titles archived: For libraries that agree to retain materials, a list or other means of identifying the titles for which that member is responsible
- Loss or damage: Level of effort expected to replace or compensate for lost or damaged items
- Financial obligations: Description of how any required fees are established. Usually the fees are not included explicitly but may be referenced in an attachment to facilitate future changes.
- Withdrawal and termination: Process for an individual member to withdraw from the agreement, and any process for the overall agreement to be terminated by the group.

For shared collections that involve long-term commitments to specific individual titles and especially in cases where funds are collected from members, the agreement almost always takes the form of a document such as an MOU that must be signed explicitly by representatives of the member institutions. In less structured cases, such as shared storage facilities where the deposited volumes become the shared copy of record, the agreement may take the form of a document adopted by the governing body and posted in a public place such as the group’s website (e.g. WRLC Shared Collection).

Current Programs

The following section provides summary details for a selection of current shared collection programs:

- ASERL Cooperative Journal Retention Program
- CRL JSTOR Print Archive Project
- CIC Shared Print Repository
- COPPUL Shared Print Archive Network (SPAN)
- Five College (MA) Library Depository
- Florida Academic Repository (FLARE)
- Midwest Collaborative for Library Services (MCLS) Shared Monographic Print Storage Project
- Orbis Cascade Alliance Distributed Print Repository
- Pennsylvania Academic Library Consortium (PALCI) Print Journal Archive
- Washington Research Library Consortium (WRLC) Shared Collection
- Western Regional Storage Trust (WEST)
<table>
<thead>
<tr>
<th>Program Name:</th>
<th>ASERL Cooperative Journal Retention Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Website:</td>
<td><a href="http://www.aserl.org/programs/cooperative-journal-retention/">http://www.aserl.org/programs/cooperative-journal-retention/</a></td>
</tr>
<tr>
<td>Members:</td>
<td>22 out of 40 ASERL libraries</td>
</tr>
<tr>
<td>Format Archived:</td>
<td>Journals</td>
</tr>
<tr>
<td>Selection:</td>
<td>Titles nominated by participating libraries.</td>
</tr>
<tr>
<td>Archive location:</td>
<td>Distributed, libraries and storage facilities</td>
</tr>
<tr>
<td>Retention Period:</td>
<td>Through December 31, 2035</td>
</tr>
<tr>
<td>Ownership:</td>
<td>Original owner</td>
</tr>
<tr>
<td>Validation:</td>
<td>Volume level, “serviceable condition”</td>
</tr>
<tr>
<td>Access/Delivery:</td>
<td>No request mechanism specified. Delivery by electronic or paper duplication, or in-house use at discretion of the owning library</td>
</tr>
<tr>
<td>Business Model:</td>
<td>Libraries absorb own expenses</td>
</tr>
<tr>
<td>Governance:</td>
<td>Steering Committee consisting of one representative of each participating library and a liaison from the ASERL Board of Directors</td>
</tr>
</tbody>
</table>
Program Name: Center for Research Libraries (CRL) JSTOR Print Archive Project

Website: [http://www.crl.edu/archiving-preservation/print-archives/crl-administered/jstor](http://www.crl.edu/archiving-preservation/print-archives/crl-administered/jstor)

Link to MOU: Not available

Members: Open to all 260 CRL members

Format Archived: Journals

Selection: By publisher or aggregator: JSTOR titles

Archive location: Centralized at CRL facility

Retention Period: Not specified

Ownership: Transfers to CRL

Validation: Issue-level

Access/Delivery: Physical volumes by ILL to the requesting library.


Governance: Existing CRL governance. No separate governance for this program.
Program Name: CIC Shared Print Repository

Website: [http://www.cic.net/Home/Projects/Library/Home.aspx](http://www.cic.net/Home/Projects/Library/Home.aspx)

Link to MOU: Not available

Members: 10 of 13 CIC members

Format Archived: Journals

Selection: By publisher: Elsevier and Wiley titles

Archive location: Centralized at Indiana University storage facility

Retention Period: 25 years, to Dec. 31, 2035

Ownership: Original owner

Validation: Volume-level

Access/Delivery: Delivery by electronic or paper duplication, or transfer to the requesting library for in-house use

Business Model: Annual member fees to cover ingest, storage, transport, project administration.

Governance: Governing Board composed of the Library Directors of the participating CIC libraries and the Director of the Center for Library Initiatives (ex officio).
Program Name: Council on Prairie and Pacific University Libraries (COPPUL) Shared Print Archive Network (SPAN)

Website: http://coppul.ca/projects/SPAN.html

Link to MOU: http://coppul.ca/projects/SPAN%20AgreementApril2012revWEB.pdf

Members: 19 of 22 COPPUL full members

Format Archived: Journals and optional less-managed process for unique monographs

Selection: Custom collection analysis, risk analysis

Archive location: Distributed, libraries and storage facilities

Retention Period: Varies. Low-Risk, until December 31, 2022 (10 years), Moderate-Risk and Higher-Risk, until December 31, 2036 (25 years).

Ownership: Archive holder (usually the original owner) or gifted to the Archive holder

Validation: Varies by risk category.

Access/Delivery: Not specified.

Business Model: Annual member fees to cover program costs including ingest and validation at member sites.

Governance: SPAN Management Committee appointed by COPPUL Board of Directors, composed of representatives from four (4) libraries participating in the Network, including representation from various sizes of libraries, more than one province, various areas of expertise (e.g. Library Director, Technical Services, Collections Management, etc.), and of Archive Holders, Builders, and Supporters.
Program Name: Five College Library Depository

Website: https://www.fivecolleges.edu/libraries/depository

Link to MOU: Not available

Members: Amherst College, Hampshire College, Mount Holyoke College, Smith College, and the University of Massachusetts, Amherst

Format Archived: Journals

Selection: Journals deposited by members become shared copies.

Archive location: Centralized at Five Colleges shared library storage facility

Retention Period: Not specified

Ownership: Transfers to the consortium (Five Colleges Inc).

Validation: None.


Business Model: Cost-sharing through the consortium membership fee.

Governance: Existing consortium governance. No separate governance for this program.
Program Name: Florida Academic Repository (FLARE)

Website: http://csul.net/node/774

Link to MOU: http://csul.net/sites/csul.fcla.edu/uploads/SUSSC_Policy_APPROVED_March_2012r.pdf

Members: 11 members of the State University System of Florida

Format Archived: Journals and monographs

Selection: Shared storage copy. Volumes deposited by members become shared copies.

Archive location: Centralized at University of Florida storage facility

Retention Period: “Permanent transfer”, “material that is donated to the facility is expected to remain in the facility”

Ownership: Transfers to the University of Florida

Validation: Volume-level review for physical condition

Access/Delivery: Digital delivery preferred. Physical volume for in-library use only.

Business Model: Annual assessments to fund operating costs of the shared collection facility.

Governance: Council of State University Libraries (CSUL)
<table>
<thead>
<tr>
<th><strong>Program Name:</strong></th>
<th>Midwest Collaborative for Library Services (MCLS) Shared Monographic Print Storage Project</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Website:</strong></td>
<td><a href="http://mcls.org/blog/?p=1697">http://mcls.org/blog/?p=1697</a></td>
</tr>
<tr>
<td><strong>Link to MOU:</strong></td>
<td>Not available. MOU being developed.</td>
</tr>
<tr>
<td><strong>Members:</strong></td>
<td>8 academic libraries out of approximately 300 MCLS members (academic and public)</td>
</tr>
<tr>
<td><strong>Format Archived:</strong></td>
<td>Monographs</td>
</tr>
<tr>
<td><strong>Selection:</strong></td>
<td>Custom analysis by Sustainable Collections Services (SCS) based on holdings overlap and circulation</td>
</tr>
<tr>
<td><strong>Archive location:</strong></td>
<td>Distributed, libraries and storage facilities</td>
</tr>
<tr>
<td><strong>Retention Period:</strong></td>
<td>15 years.</td>
</tr>
<tr>
<td><strong>Ownership:</strong></td>
<td>Original owner</td>
</tr>
<tr>
<td><strong>Validation:</strong></td>
<td>Volume-level review for physical condition.</td>
</tr>
<tr>
<td><strong>Access/Delivery:</strong></td>
<td>Digital delivery preferred. Physical volume for in-library use only.</td>
</tr>
<tr>
<td><strong>Business Model:</strong></td>
<td>Libraries absorb own expenses</td>
</tr>
<tr>
<td><strong>Governance:</strong></td>
<td>Not finalized</td>
</tr>
</tbody>
</table>
Program Name:  Orbis Cascade Alliance Distributed Print Repository (DPR)

Website:  http://www.orbiscascade.org/index/cdmc-current-work


Members:  All 36 members of the Orbis Cascade Alliance

Format Archived:  Journals

Selection:  By publisher:  JSTOR Arts & Sciences I and II and American Chemical Society journals

Archive location:  Distributed, libraries and storage facilities

Retention Period:  25 years.

Ownership:  Archive holder (usually the original owner) or gifted to the Archive holder

Validation:  Volume-level review for physical condition.

Access/Delivery:  Digital delivery preferred.  Physical volume for in-library use only.

Business Model:  Libraries absorb own expenses

Governance:  Existing consortium governance.  No separate governance for this program.
<table>
<thead>
<tr>
<th><strong>Program Name:</strong></th>
<th>Pennsylvania Academic Library Consortium (PALCI) Print Journal Archive</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Website:</strong></td>
<td><a href="http://www.palci.org/collection-management/">http://www.palci.org/collection-management/</a> (scroll down)</td>
</tr>
<tr>
<td><strong>Link to MOU:</strong></td>
<td>Not available.</td>
</tr>
<tr>
<td><strong>Members:</strong></td>
<td>About 10 of PALCI’s 71 members</td>
</tr>
<tr>
<td><strong>Format Archived:</strong></td>
<td>Journals</td>
</tr>
<tr>
<td><strong>Selection:</strong></td>
<td>By publisher: journals published by the American Chemical Society (ACS), the American Institute of Physics (AIP), and the American Physical Society (APS).</td>
</tr>
<tr>
<td><strong>Archive location:</strong></td>
<td>Distributed, libraries and storage facilities. Penn State keeping a dark archive copy, others are light archives.</td>
</tr>
<tr>
<td><strong>Retention Period:</strong></td>
<td>10 years, to Dec. 31, 2019.</td>
</tr>
<tr>
<td><strong>Ownership:</strong></td>
<td>Archive holder (usually the original owner) or gifted to the Archive holder</td>
</tr>
<tr>
<td><strong>Validation:</strong></td>
<td>Issue-level review for physical condition and completeness.</td>
</tr>
<tr>
<td><strong>Access/Delivery:</strong></td>
<td>Digital delivery preferred. Physical volume for in-library use only.</td>
</tr>
<tr>
<td><strong>Business Model:</strong></td>
<td>Libraries absorb own expenses</td>
</tr>
<tr>
<td><strong>Governance:</strong></td>
<td>Existing consortium governance. No separate governance for this program.</td>
</tr>
</tbody>
</table>
Program Name: Washington Research Library Consortium (WRLC) Shared Collection

Website: [http://www.wrlc.org/offsite/](http://www.wrlc.org/offsite/)

Link to MOU: [http://www.wrlc.org/offsite/storpolicies.html](http://www.wrlc.org/offsite/storpolicies.html)

Members: All 9 WRLC members

Format Archived: Journals and monographs

Selection: Shared storage copy. Volumes deposited by members become shared copies.

Archive location: Centralized at WRLC storage facility

Retention Period: Permanent

Ownership: Original owner

Validation: Minimal review for physical condition

Access/Delivery: Digital delivery preferred. Physical volume for in-library use only.

Business Model: Cost-sharing through the consortium membership fee

Governance: Existing consortium governance. No separate governance for this program
**Program Name:** Western Regional Storage Trust (WEST)

**Website:** [http://www.cdlib.org/west/](http://www.cdlib.org/west/)

**Link to MOU:** [http://www.cdlib.org/services/west/docs/WESTProgramStatement.pdf](http://www.cdlib.org/services/west/docs/WESTProgramStatement.pdf)

**Members:** 103 libraries including 3 sets of consortial memberships

**Format Archived:** Journals

**Selection:** Custom collection analysis, risk analysis

**Archive location:** Distributed, libraries and storage facilities

**Retention Period:** 25 years, until Dec. 31, 2035

**Ownership:** Archive holder (usually the original owner) or gifted to the Archive holder

**Validation:** Varies by risk category. See validation standards at [http://www.cdlib.org/services/west/docs/WESTStandards_Issue_VolumeLevelValidation.docx](http://www.cdlib.org/services/west/docs/WESTStandards_Issue_VolumeLevelValidation.docx)

**Access/Delivery:** Digital delivery preferred. Physical volume for in-library use only.

**Business Model:** Annual member fees to cover program costs including ingest and validation at member sites.

**Governance:** WEST Executive Committee, composed of representatives from nine (9) WEST members (with representatives for Archive Holders, Archive Builders and Non-Archive Holders), elected by the full WEST membership.
ReCAP Discovery to Delivery Project
Description and Analysis of Current Workflows

I. Introduction
In March 2012, Ruth Fischer and Rick Lugg of Sustainable Collection Services (SCS) were engaged by NYPL to serve as the Workflow Consultant on the ReCAP Discovery and Delivery Project, reporting to the Workflow/Technology Committee. This narrative and the accompanying grid comprise a functional comparison of the ReCAP-related workflows currently in place at New York Public Library (NYPL), Princeton University Libraries (PUL), and Columbia University Libraries (CUL). This includes:

- ‘Pre-facility’ tasks (Identifying and processing materials for storage)
- ‘In-facility’ tasks (Accessioning, inventory management, picking & refilling at ReCAP)
- ‘Post-facility’ tasks (Processing user requests, notifications, receipt & return of requested items)

This first deliverable is intended to be descriptive rather than prescriptive. In our experience, it is important to understand current workflow practice before suggesting changes. This is especially important in this instance, which involves four complex operations that have evolved independently. This report will serve as the foundation for the second SCS deliverable: a series of recommendations for adapting workflows to support a “shared ReCAP collection”. Those recommendations will be submitted in September, and will be based on reconciliation of our work with that of the Technology Consultant and the Planning Consultant.

In April 2012, SCS visited the ReCAP facility and each of the three libraries, to observe various aspects of current workflows and to interview key personnel in each location. These visits were preceded by a questionnaire, a request for procedural and policy documents, and multiple individual and conference calls. Overall, SCS interviewed approximately 20 people and gathered hundreds of pages of related documentation. To a person, those we met were fully engaged in our discovery process and generous with their time and feedback. We thank them for their patience and forbearance as we verified details.

In June 2012, SCS compiled a four-column workflow grid, which outlined ReCAP workflows of the three partner libraries, as well as the steps performed at ReCAP itself. This grid was reviewed and corrected by knowledgeable staff and administrators at all four locations. The full comparison accompanies this report as a separate spreadsheet, and serves as the basis of our analysis. Although each accomplishes a version of the same basic tasks, these workflows are complex and subject to a good deal of variability. This shouldn’t be surprising, since they were designed by each library independently, in the context of its individual practices and systems. In short, these workflows were not designed with sharing in mind, and we should remember that as we wrestle with the variations and complexity. Although SCS, the libraries and ReCAP have all worked hard to assure accuracy, corrections and/or clarifications are probably needed and are still most welcome.
Our understanding of project goals is as follows:

- To identify those materials currently housed in ReCAP that could be considered ‘shareable’. As noted in Lizanne’s Planning Report, this ‘shareable’ collection is understood to be a subset of what is currently housed there or may be transferred there in future.

- To ensure availability of real-time item status information on shared ReCAP holdings – usable by all three library systems and discovery layers.

- To reduce or eliminate the accessioning of any “new duplicates” into the ReCAP facility. That is, to design workflows such that none of the libraries will send volumes to ReCAP that are already held there.

- To consistently achieve promised delivery times for patrons at all three institutions (no degradation of service) regardless of which library owns the requested material.

With regard to this final point, SCS is compelled to note that administrators at two of the three partner libraries find current service levels to be in need of improvement, especially since increasingly higher-use materials may be housed there. Although it is not explicitly a goal of this project, SCS recommends establishing an expectation for improved delivery times as new workflows are designed.

This report summarizes our current understanding of existing circumstances, and poses number of questions for project managers. Some of these will need to be answered before new workflows can be effectively designed. In addition to the separate workflow grid, our observations are organized into six short sections, as follows:

II. Notable policy differences (p. 3)
III. Notable workflow differences (p. 5)
IV. Theoretical best-case and worst-case user experiences (p. 8)
V. Potentially ‘shareable’ ReCAP customer codes (p. 10)
VI. GFA/LAS workflow related functions and features (p. 11)
VII. Workflow support for a shared ReCAP collection
II. Notable policy differences
The first two sections of the workflow grid highlight various attributes of the three libraries that have bearing on this project. Again, we’ve done our best to accurately record details, and key contacts in each library have reviewed this information, but there may still be errors or omissions. We welcome corrections. Following are some observations derived from the grid, organized along the same lines.

Circulation policies
Perhaps the most significant policy difference among the three libraries relates to circulation. While Columbia and Princeton allow their unrestricted materials to be checked out to users, NYPL policy stipulates that research materials are for in-library use only. (If we understand correctly, the one systematic exception to this rule is for MaRLI patrons. That is, Columbia and NYU patrons with a MaRLI card are allowed to take NYPL books out of the library.) In order to assure consistency in both the user experience and materials handling, some adjustment or policy reconciliation may be necessary.

For example, to maximize speed of delivery to the patron, ‘shared’ ReCAP materials will be delivered to the location specified in the user’s request. If a Columbia patron requests an item from the ReCAP shared collection that is owned by NYPL, it would be delivered to a Columbia location. But whose circulation rules would apply? Would that item be checked out to the user (Columbia’s policy) or would it be offered for in-house use only (NYPL’s policy)? The reverse question also arises. If a Columbia item is requested by an NYPL researcher, will that researcher be required to use it in a reading room, or will that patron have the privilege of taking it home?

Delivery benchmarks
Each of the three libraries now has its own delivery benchmark:

- Columbia University Libraries advertise delivery of ReCAP requests "within two business days". The ‘business’ day is defined according to the last file transfer time (2:45 pm, Monday through Friday). In this model, 48% of requests are filled in one business day and the rest are filled in two business days. This is considered satisfactory.

- The expectation at Princeton is that ReCAP requests are ready for patron pick-up by 5:00 pm on the business day following placement of the request. This is of prime importance to Princeton stakeholders. 90% of requests are currently delivered according to this specification. The remaining 10% take an additional day, which at present is not accounted for. It is of utmost importance to eliminate this 10% and achieve 100% next day delivery.

- At NYPL, all weekday requests made by 2:30 pm arrive the next day. Given the high level of public attention being paid the library renovation and related collection moves, it is very important to administrators that service improvements be realized as part of the next phase of the ReCAP
One aspect of improved service will be to ensure 24 hour delivery of all requested items. This implies a 7-day week, though that needs to be confirmed.

Since coordination of individual ReCAP workflows has not been necessary to date, it is not surprising that delivery expectations differ and are expressed differently. In future, if some collections and related workflows are to be shared, more consistency will likely be needed.

**Consortial/Cooperative affiliations**
As depicted on the second page of the workflow grid, each library has its own unique set of borrowing partners. Or, as Lizanne Payne puts it in her Planning Report, “there is no common resource-sharing consortium among ReCAP partners.” SCS understands that the scope of this project does not include the development of any new sharing agreements. We interpret this to mean that when (for example), a Rapid ILL request is received by Columbia, the request will only be forwarded to ReCAP if the “shared ReCAP copy” belongs to Columbia. If this is correct, we should seek to confirm that discoverability of the “shared collection” will be restricted to Columbia, Princeton, and NYPL patrons. Perhaps this is already the expectation.

**Fair Use**
During the course of our interviews, SCS was informed that the fair use guidelines at ReCAP are not entirely consistent with fair use guidelines followed at the campuses of the partner institutions. The partners will need to agree on guidelines for fair use that will satisfy patrons’ needs as well as the desire of all three institutions’ offices of general counsel.
III. Notable workflow differences

Based on our understanding of current workflows, there are several notable differences among the three libraries, both for accessioning materials to ReCAP and retrieval of items for patrons. The workflow at the ReCAP facility itself does not vary, other than the use of different couriers to move materials. SCS encourages readers to spend some time reviewing the workflow grid that accompanies this report, and from which these highlights are drawn. Notable differences are summarized below. Please note that SCS makes no judgment on these workflows. They were designed to support local needs, not shared processes. While some will affect a common workflow more than others—and may therefore ultimately need to change—they have served well for many years.

Selection for off-site storage/De-duplication

Item- or title-specific decisions are made by selectors at Columbia and Princeton, aided by system-generated candidate lists. Both libraries attempt to prevent duplicates of circulating copies from their own collections from being sent to ReCAP; these are instead withdrawn. At present, no consideration is given to potential duplication with titles from the other ReCAP libraries.

At NYPL, selection for off-site storage currently operates at massive scale. Decisions involve entire ranges or locations. These are identified and prioritized for relocation by the Collection Strategy Department. Duplicates within the NYPL collections are sent off-site along with other materials. No titles are withdrawn. Until the new ReCAP modules are ready, NYPL materials are being sent to the Tri-States Depository in upstate New York, which is variously referred to as TSD, Clancy, Brewster, or Patterson.

Physical preparation

Columbia uses “smart” barcodes for many (but not all) ReCAP materials. The ReCAP customer code doubles as the barcode prefix. Each code has a corresponding color strip. Princeton and NYPL use standard barcodes.

Princeton uses Mylar bags for fragile and multipart items, whereas Columbia and NYPL use Tyvek.

At Columbia the call number is transcribed onto the verso of the title page, preceded by the word “Offsite”. At NYPL, special ReCAP call numbers (format 12-xxx) are written inside the book and printed on spine labels. At Princeton, no manual transcription occurs.

Princeton places a pink ReCAP sticker on the spine prior to initial accessioning. Columbia places a pink ReCAP sticker on the spine when the item is retrieved for a patron. NYPL places a yellow ReCAP sticker on the spine when the item is retrieved for a patron. (NYPL/Brewster materials have a pink sticker.)

Pre-accessioning record maintenance for ReCAP accessions
Princeton changes the Voyager location code to one that corresponds to the ReCAP customer code. This is the mechanism that allows the offsite request form as well as appropriate borrowing rules to display in the catalog. Princeton’s materials continue to have an “available” status throughout the transport and accessioning processes.

At NYPL, each item is checked-out to a pseudo patron which changes the status to “unavailable”.

At Columbia, the CLIO location code is changed to a transitional location (like glx4off) for small-scale recon projects and for processing new acquisitions directly to ReCAP. For large scale projects (like the current 12th Tier move) location codes remain unchanged until after accessioning (described below). If we understand correctly, Columbia’s materials continue to have an “available” status throughout the transport and ReCAP accessioning processes.

**Pick-up schedule for new ReCAP accessions**
Columbia is currently sending batches of 4,500 items 3-5 times per month. Because these deliveries arrive in such large volume, there can be up to a week of lag time before materials are sized at ReCAP.

Princeton is currently sending batches of up to 1,500 items per day.

NYPL is currently sending batches of about 65,000 items per week (to TSD/Brewster).

As materials arrive at ReCAP, they are sized, verified, and put into trays. Accession reports are generated daily and made available to the libraries for retrieval.

**Post-accessioning record maintenance in library catalog**
At Columbia, lists of newly accessioned barcodes (generated daily by ReCAP) are retrieved on a weekly basis. In a batch process, the location code in Voyager is updated to one of many offsite locations (off, glx). This causes the “Request Button” to appear in the CLIO (Columbia’s catalog) and in Blacklight (its discovery layer).

At NYPL, lists of items accessioned (currently at Brewster) are used to drive a daily batch check-in process in the library’s Millennium system. This changes the status back to “available” and updates the location, the OPAC message, and the Item Agency fields. The location reflects both its current offsite location AND the originating library building/collection; the Agency field reflects the ReCAP customer code; and the OPAC message = “ADV REQUEST”.

No post-accessioning record maintenance is performed at Princeton. (The Voyager location code was changed to the ReCAP customer code before items were sent.)

**Discovery**
Except for system/platform differences, discovery processes are similar in all three libraries.
**Patron request of ReCAP item**

All three libraries have a locally-developed request form for retrieval of items from ReCAP linked to their discovery tools. (At NYPL the form is only implemented in their WebPac – to be integrated in BiblioCommons within the next 6-8 months.) These forms display automatically when an item is held offsite, triggered by various versions of offsite location codes.

Columbia maintains a separate local indicator of status for items that have been requested from ReCAP but not yet retrieved. Among other purposes, this enables Columbia to prevent display of an item that is currently checked out to another user, or is in some stage of processing. This is understood to be a cumbersome workaround, designed to compensate for the lack of real-time status information from ReCAP’s GFA/LAS system. Princeton and NYPL have elected to live with the information gap.

**Requests transmitted to ReCAP**

All three libraries collect requests into batches that are sent to ReCAP in a specified format 3 times every weekday. Columbia and Princeton transfer requests at 7:15 am, 11:45 am, and 2:45 pm. NYPL transfers ReCAP requests at 8 am, 12 pm, and 2:30 pm. In a separate data flow, NYPL also transfers requests to TSD/Brewster at 1 and 5 pm.

Requests that error out (and cannot be filled) are communicated to patrons more or less manually at all three libraries.

‘In Process’ notifications

Some additional clarification is needed in this area. For NYPL, it is reported that an in-process delivery notice is generated by ReCAP and delivered to the patron. As far as we can tell, ReCAP is not generating such notices for Princeton or for Columbia patrons—and it seems unlikely that ReCAP would have the patron’s email address. Therefore, we think that this in-process notice must actually be generated from Millennium upon NYPL’s import of the ReCAP picking report?

Princeton’s system shows as ‘available’ throughout the request, retrieval, and delivery process. Is there mechanism for notifying patrons of their request’s status?

Columbia?

So we need to find out more about how, when, and if this step occurs at each institution. We also need to determine how important this may be, given that most requests are filled within 24-48 hours—is an interim status report really needed, other than for unfillable requests?

**Transport of ReCAP requests**

Columbia’s requests are picked up at ReCAP at 4:00 AM (Monday – Friday) and delivered to the central library before opening.
NYPL’s requests are picked up at ReCAP at 4:00 AM (Tuesday – Saturday) and delivered to the central library before opening – routed to other research locations by 9:00 AM.

Princeton ReCAP requests are picked up at 12:30 pm (Monday – Friday) and delivered to all campus stops, reaching the central library last (by 2:00 PM).

**Patron notification upon delivery**
Princeton and NYPL use the circulation ‘hold’ feature of their respective Voyager systems to generate an email notification telling the patron that his/her request is available. SCS notes that each library’s schedule for generating patron notifications is different, and that the timing of notice generation can significantly impact the patron’s experience of delivery time. In addition, the notice generation cycles do not necessarily coincide with the arrival of material at the library from ReCAP.

Columbia uses their in-house system rus to generate an email to the patron.

**Patron pick-up at library**
All three libraries use their ILS circulation modules to charge ReCAP materials out to patrons. As noted previously, NYPL items are used only in designated research locations, while Columbia’s and Princeton’s circulate outside the library.

**ReCAP items returned to library**
At Columbia and Princeton, items are discharged manually by circulation staff. These items are not re-requestable until they are re-filed at ReCAP.

At NYPL, the status is set to “in transit” at the time of the return, but check-in (discharge) is accomplished in batch the next day prompted by the re-filing report generated by ReCAP. During this batch check-in, the status is set back to “available.”

These examples serve to illustrate a few of the differences in timing and procedures among the three libraries. Some differences clearly matter more than others, but overall it seems obvious that some reconciliation and a greater degree of consistency will be needed to support a shared ReCAP collection in a manner that maintains or improves service to users. As mentioned in the introduction, SCS recommendations related to changed/enhanced workflows will be the focus of our September report.
IV. Theoretical best-case and worst-case user experiences

Given the differences in practice and policy just outlined, it’s not surprising that patron experiences may vary. At present, however, patrons remain largely unaware of any variation because each patron works through a single institution. The most critical aspects of good customer service include consistency and predictability. At present, SCS suggests that many patrons would find it difficult to predict with any certainty when their request would be fulfilled, even when working through a single institution.

It may also be helpful to consider how much research is conducted outside of standard business hours, especially now that Web-based request forms are in place at all three libraries. SCS recommends additional investigation into the timing of ReCAP requests at each library: when are requests submitted? It might be quite informative to examine the pattern of requests across the day. From a workflow perspective, user requests are the starting point, and ideally their frequency pattern and volume would shape the supporting processes. This may or may not be fully practicable, but we believe the data should be collected and examined. From the user’s point of view, the fulfillment clock starts at the moment an item is requested.

Because transmission of requests, confirmations, and deliveries from ReCAP occur on relatively fixed schedules, the period between request and delivery is the most predictable. Research at Columbia suggests that 48% of ReCAP requests are available to the requestor on the following business day. This implies that just 48% of requests are submitted before 2:45 pm, Monday through Friday. As we see it, information like this about specific user behaviors should be carefully incorporated into the design of new workflows, staffing models, and possibly hours of operation.

Given the fact that Columbia and NYPL share the same transport company and delivery schedule, and despite the operational differences noted above, delivery timeframes are very similar. The best-case scenario for NYPL and for Columbia patrons is to be notified that their material is available approximately 24 hours after the request was submitted. (These would be items requested between 8:00 AM and 2:45 PM on Day One, transmitted to ReCAP by 3:00, picked by end of day, picked up at 4:00 AM Day Two, delivered to the library by 8:00, with patron notification by 12:00 PM.) Again, this occurs about 48% of the time for Columbia patrons. We do not know the best-case percentage for NYPL.

For Princeton, the best-case scenario is somewhat better, due to proximity to the ReCAP facility and Princeton’s mid-day pick-up. In an absolute best case, a Princeton patron is notified of availability 10-16 hours after request submission—though that is only if their request submission occurs between 1:00 AM and 7:00 AM. (These are captured in the 7:00 AM transmission and can be ready for the 12:30 PM pick-up.) While this may seem unlikely to some of us, it’s conceivable that this may be common at some points in the semester. Here again, we don’t know the percentage that fit this scenario, but this could be learned in quantifying activity around the clock. The following chart details the three theoretical best-case scenarios.

ReCAP: Current Workflows
<table>
<thead>
<tr>
<th>Theoretical Best-Case Scenarios</th>
<th>Columbia</th>
<th>Princeton</th>
<th>NYPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day and time of request</td>
<td>8 am – 2:45 pm Monday - Friday</td>
<td>1 am -7 am Monday - Friday</td>
<td>9 am – 2:30 pm Monday – Friday</td>
</tr>
<tr>
<td>Day and time of delivery to the library</td>
<td>Before 8 am on the following day (Monday – Friday)</td>
<td>At 2:00 pm on the same day (Monday – Friday)</td>
<td>Before 9 am on the following day (Tuesday – Saturday)</td>
</tr>
<tr>
<td>Day and time of patron notification</td>
<td>Before noon on the day following the request <strong>22 - 26 hours after the request was submitted</strong></td>
<td>By 5:00 pm on the day requested – 5 pm is when the Voyager ‘hold’ notices are generated <strong>10 – 16 hours after the request was submitted</strong></td>
<td>Before noon on the day following the request <strong>22 – 26 hours after the request was submitted</strong></td>
</tr>
</tbody>
</table>

Worst-case scenarios (exclusive of errors) are actually quite similar across the group, insofar as Friday afternoon requests (those submitted after 2:30 or 2:45) will not be delivered to the library until the following Monday afternoon (Princeton) or Tuesday morning (NYPL and Columbia). The following chart details the three theoretical worst-case scenarios.
# Theoretical Worst-Case Scenarios

<table>
<thead>
<tr>
<th></th>
<th>Columbia</th>
<th>Princeton</th>
<th>NYPL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Day and time of request</strong></td>
<td>2:46 pm (or later) on Friday afternoon</td>
<td>2:46 pm (or later) on Friday afternoon</td>
<td>2:31 pm (or later) on Friday afternoon</td>
</tr>
<tr>
<td><strong>Day and time of delivery to the library</strong></td>
<td>Before 8 am on the following Tuesday morning.</td>
<td>At 2 pm on the following Monday.</td>
<td>Before 9 am on the following Tuesday morning.</td>
</tr>
<tr>
<td><strong>Day and time of patron notification</strong></td>
<td>Before noon on the following Tuesday 69 hours after the request was submitted Note that during the initial accessioning process, staff expect a 2-4 week gap between physical transfer and requestability in the OPAC.</td>
<td>By 5:00 pm on Monday 72 hours after the request was submitted Note that Mondays with exceptionally large numbers of ReCAP receipts can sometimes roll into Tuesday for patron notification.</td>
<td>Before noon on the following Tuesday 68 hours after the request was submitted</td>
</tr>
</tbody>
</table>

While best and worst cases can help set the parameters for delivery, they do not reflect typical service. But they help make our overall point: the patron experience of timeliness depends on a combination of:

- Time of request
- Request transmission cycles
- ReCAP picking cycles / hours of operation
- Delivery cycles
- Check-in workload at receiving library
- Patron notification cycles

No single entity controls (or can control) the process from beginning to end. To achieve more consistent delivery within specified windows will require coordination across all of these functions.
V. Potentially ‘shareable’ ReCAP customer codes

In contemplating workflows for a shared ReCAP collection, we thought it would be useful to gauge the approximate size of that collection—or at least its initial size. The agreed definition, drawn from Lizanne Payne’s Planning Report: “materials housed at the ReCAP facility (or transferred there in the future) that meet certain selection criteria and are placed under a retention agreement or joint ownership agreement.” While, in the categories made immortal by Donald Rumsfeld, there are several “known unknowns” in this definition, we can look at materials housed in the facility now. As a starting point, SCS identified the existing customer codes that could conceivably be ‘shared’—i.e., those that are currently unrestricted. We then tallied the total number of items accessioned by ReCAP into these codes to date. As an estimate of likely transaction activity, we incorporated the number of items retrieved during the last fiscal year (7/1/10 – 6/30/11). The table below gives us a preliminary view of the workload that may be involved in creating and supporting a shared collection.

Note that these numbers do not count titles with unrestricted customer codes that are currently awaiting transfer to ReCAP, the biggest group of which is the NYPL titles now going to TSD/Brewster. Nor does this include a count of the retrievals from Brewster. Note also that while all of these customer codes are listed on the ReCAP website with no request or delivery restrictions, four have no holdings either. As we understand it, those with 0 holdings are simply stop (or delivery) codes.

<table>
<thead>
<tr>
<th>Customer Code</th>
<th>Customer Name</th>
<th>Total items</th>
<th>Retrieved during the last fiscal year</th>
</tr>
</thead>
<tbody>
<tr>
<td>CJ</td>
<td>Journalism Library (Columbia)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CU</td>
<td>Columbia Standard (Columbia)</td>
<td>2,576,788</td>
<td>45,196</td>
</tr>
<tr>
<td>EV</td>
<td>East Asian Vernacular (Columbia)</td>
<td>336,581</td>
<td>7,919</td>
</tr>
<tr>
<td>GC</td>
<td>Government Documents (Columbia)</td>
<td>33,422</td>
<td>264</td>
</tr>
<tr>
<td>HS</td>
<td>Health Science Library (Columbia)</td>
<td>48,447</td>
<td>700</td>
</tr>
<tr>
<td>JC</td>
<td>JSTOR Standard (Columbia)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SW</td>
<td>Social Work Library (Columbia)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PA</td>
<td>Unrestricted (Princeton)</td>
<td>1,807,683</td>
<td>19,479</td>
</tr>
<tr>
<td>QK</td>
<td>Mendel Sound &amp; Video (Princeton)</td>
<td>32,093</td>
<td>219</td>
</tr>
<tr>
<td>GP</td>
<td>Government Documents (Princeton)</td>
<td>18,183</td>
<td>273 (?)</td>
</tr>
<tr>
<td>JP</td>
<td>JSTOR Standard (Princeton)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NA</td>
<td>NYPL Standard</td>
<td>2,677,484</td>
<td>27,736</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td></td>
<td><strong>7,512,498</strong></td>
<td><strong>102,665</strong></td>
</tr>
</tbody>
</table>

A few observations:

These **7,512,498** items represent the full universe of items that could be shared given current definitions and current ReCAP holdings. Additional selection criteria are likely to be applied to identify a subset for sharing.
These 7,512,498 potentially shareable items represent approximately 78% of all items currently stored at ReCAP. During the last few years, between 1% and 2% of items accessioned under these customer codes have been retrieved annually. That translates to approximately 8,500 per month or 425 per week day.

If predictions are accurate about anticipated transfers to ReCAP, the size of the shared collection will increase by some 2 million volumes as soon as the NYPL Brewster materials are transferred in, and growth of the shared collection will likely continue at a pace of more than 500,000 volumes per year. That suggests an eventual shared collection of 9-10 million items, and an annual retrieval volume of 180,000-200,000 (using a 2% rate).

These are raw numbers, of course, and do not reflect the “certain selection criteria” that will be set. They may not fully reflect future retrieval levels either, which seem likely to increase (especially for NYPL), as a greater proportion of the print collections are stored at ReCAP. But this does give us an idea of potential overall size and activity, and suggests the scale of operation that may be needed to support the shared collection.
VI. GFA/LAS workflow support
As the core system at the ReCAP facility, Generation Fifth’s Library and Archive System (LAS) affects every aspect of those workflows. When we begin to consider possible adjustments to workflows in support of a shared ReCAP collection, it will be important to understand LAS functionality, limitations, and potential for modification. While interoperability with the respective discovery layers and library management systems is covered in the Technology Consultant’s report, SCS ‘s observations relate to the relationship between LAS and work processes. The following factors seem most likely to influence workflow redesign, and we will continue to explore the related possibilities and limitations as we prepare recommendations for our September report.

Customer codes
- The customer code, more than anything else, serves to communicate ownership of the volume.
- Customer codes also serve as the primary payment construct.
- Customer codes are directly associated with delivery locations.
- In the case of ReCAP, customer codes are also associated with specific circulation rules/parameters. These associations are apparently not typical in other LAS implementations, and, as we understand it, were established against the recommendation of GFA.
- Trays and shelves are always associated with a single customer code.
- There is no way to change a customer code in LAS. New permissions to support batch changes could potentially be developed but at present, items must be withdrawn and re-accessioned under a different code.
- Changing customer codes in batch is something that can be done only by GFA staff and even then, only if all items bearing that customer code are changed. That is, items within customer codes cannot be changed selectively.

It may be possible to accession entire trays into ReCAP. This is potentially good news for the NYPL materials currently stored at TSD/Brewster. Rather than accessioning individual items, GFA reports that transfer to ReCAP could be handled by tray (with some modification to LAS). This strategy could reduce the transfer costs significantly, and GFA has enabled this approach in other facilities. This will work only if the tray numbers assigned in TSD/Brewster do not overlap with those in use at ReCAP. NYPL is aware of this feature and has chosen customer codes (and tray numbers?) accordingly. See additional notes on this topic in the workflow grid.

It is practically impossible to fill ‘holes’ in ReCAP trays. While a standard LAS ‘location analysis report’ provides information on partially filled shelves, it cannot be used to identify partially-filled trays. This is primarily because there are no depth measurements taken on accessioned items. So when an item is removed from a tray, there is no systematic way to estimate how much space was freed. This makes it impossible for the system to assign another item to that tray. This makes de-duplication of items already in ReCAP unattractive and probably not cost-effective, since there is no good way to fill the spaces created by de-duplication. One possible exception: contiguous journal runs. Also, LAS provides reports can identify space created by ‘permanent withdrawals’ – the utility of this should be explored further.
**ReCAP picklists can be customer-code specific or not.** It will be critical to understand the operational effects of changes to customer codes. Whether the ReCAP shared collection is eventually based on a single shared customer code or an amalgamation of the current unrestricted customer codes, the ultimate decision will affect how picklists are generated and organized. It is useful to know that LAS has some flexibility in this regard. Especially if retrieval volume increases, it is essential that picking and re-filing are not negatively impacted by any changes in the customer codes.
VII. Workflow support for a Shared ReCAP Collection: Topics for final report
While this report only hints at potential changes or improvements, SCS has ideas on several topics, which will be fully developed in our September report. Some of our ideas are summarized here:

**No new duplicates at ReCAP:** Given the difficulty of removing duplicates once they are in the ReCAP facility, a goal of preventing new duplicates being accessioned seems wise. But this raises some difficult questions. How to define a duplicate? Is a duplicate to be defined only within the ‘shared ReCAP collection’? Are ‘non-shareable’ collections exempt? What will this require in terms of workflow adjustments? Will the NYPL Deaccessioning policy be reconsidered as part of this project—i.e., if NYPL chooses not to de-accession anything, how does that affect this policy?

**How to manage ReCAP customer codes:** Conversations about defining and servicing the ReCAP shared collection have already begun. One approach might be to merge all ‘sharable’ customer codes into a single new ReCAP shared collection code, supporting all relevant delivery locations. This strategy raises questions about cost accounting and customer code maintenance. Another strategy might be to retain existing codes but to assign new attributes (delivery locations, etc.) to those that are intended for sharing. Customer code maintenance would still be required, but de-accessioning and re-accessioning would not be necessary. These and other strategies will be explored.

**Higher-use items in ReCAP:** Over time, increasingly higher use items have been sent to ReCAP. These include newly acquired monographs, for which it is difficult to predict use. At Columbia, request rates measured by LC class indicate that recent requests in Science and Technology far outstrip other classes. Is there benefit in exploring this further? For NYPL, higher request rates are expected as a greater portion of the collections reside offsite. What are the workflow implications of higher levels of recall activity over the next 5–10 years? In some respects, higher transaction volume may reduce the cost-per-volume, as fixed costs such as transportation are amortized over more items. It may be useful to model scenarios at transaction volumes of 1%-5%.

**Improved delivery to patrons:** Although the workflow charge for this project is to consistently reach promised service levels, NYPL’s situation suggests that consistent 24-hour delivery may warrant exploration. Further, it may be worth considering expansion to six or even seven-days-a-week service, to help patrons accept the idea that the majority of collections are stored offsite. Cost considerations may prevent the most expansive version of this, but at minimum we will examine the interaction of request, picking, delivery, and notification cycles. And, if available, it would be useful to view the distribution of patron requests across the day/week.

**Newly acquired items going directly to ReCAP:** At present, newly acquired materials cannot be accessioned directly at ReCAP, but must first be “received” into the ILS by the owning library. Should libraries consider “receiving” materials via electronic invoice ---and allowing the books to be shipped
directly to ReCAP? What are the possible risks and/or rewards? To what degree are these likely to be part of the shared ReCAP collection?
### ReCAP: Current Workflows (July 2012)

<table>
<thead>
<tr>
<th>Collections</th>
<th>Columbia</th>
<th>Princeton</th>
<th>NYPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items at ReCAP</td>
<td>3.7 million</td>
<td>2.1 million</td>
<td>3.5 million</td>
</tr>
<tr>
<td>Potentially 'shareable' customer codes</td>
<td>CJ, CU, EV, GC, HS, JC, SW (50 customer codes in all)</td>
<td>PA, QK, GP, JP (30 customer codes in all)</td>
<td>NA (30 customer codes at present)</td>
</tr>
<tr>
<td>Anticipated transfers to ReCAP</td>
<td>175,000/year. Columbia is actually holding off sending some stuff (like large business serials), assuming that they will duplicate the NYPL collections destined for ReCAP. These materials are currently staged in the Lehman Social Sciences Library. Materials are also staged in the Butler basement as part of the routine ReCAP workflows.</td>
<td>200,000/year through 2020, of which, ca. 140,000 eclectically selected from campus collections and ca. 60,000 from local remote books storage, the Forrestal Annex. Forrestal Annex contains ca. 500,000 volumes, all of which are targeted for transfer to ReCAP, but as a secondary priority to campus collection space management needs.</td>
<td>2.3 million asap - many of these are being transported to the Tri-State Depository (also referred to as TSD, Brewster Facility, and/or Patterson facility) until the new ReCAP modules are completed.</td>
</tr>
<tr>
<td>Volumes in various campus locations</td>
<td>4 million</td>
<td>5 million</td>
<td>The goal is to keep 1.5 million books onsite in BPSE. Special Collections will stay onsite.</td>
</tr>
<tr>
<td>Circulating versus non-circulating</td>
<td>Columbia</td>
<td>Princeton</td>
<td>NYPL</td>
</tr>
<tr>
<td>How to reconcile these differences with regard to a &quot;shared&quot; collection?</td>
<td>For the vast majority of resources, users are allowed to check them out. Columbia offers a Faculty Document Delivery Service that allows faculty affiliated with the Morningside Heights Campus, Barnard College, and Union Theological Seminary to request Butler resources to be delivered electronically.</td>
<td>Users are allowed to check out the vast majority of resources.</td>
<td>In general, research collection materials do not circulate; are used in its reading rooms only. One current exception relates to MaRLI patrons. Perhaps this exception would be extended to Princeton patrons?</td>
</tr>
<tr>
<td>Delivery benchmarks</td>
<td>Columbia</td>
<td>Princeton</td>
<td>NYPL</td>
</tr>
<tr>
<td>What are the requirements/expectations for the request to delivery time frame?</td>
<td>CUL advertises delivery &quot;within two business days&quot;. ReCAP's &quot;business&quot; day is defined according to the file transfer times. Business day is 6 am - 2:45 pm, Monday through Friday. In this model, 48% of requests are filled in one business day and the rest are filled in two business days. This is considered satisfactory.</td>
<td>The clear and firm expectation is that requests are ready for patron pick-up by 5:00 pm on the business day following placement of the request. This is of prime importance to Princeton stakeholders. 90% of requests are currently delivered according to this specification. The remaining 10% take an additional day, which at present is not accounted for. It is of utmost importance to eliminate this 10% and achieve 100% next day delivery.</td>
<td>All weekday requests made by 2:30 pm will arrive the next day. Given the very high level of public attention being paid the library renovation, it is very important to administrators that service improvements be realized as part of the next phase of the ReCAP project. One aspect of improved service will be to ensure 24 hour delivery of all requested items.</td>
</tr>
</tbody>
</table>
# ReCAP: Current Workflows (July 2012)

<table>
<thead>
<tr>
<th>Installed systems</th>
<th>Columbia</th>
<th>Princeton</th>
<th>NYPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated Library System</td>
<td>Voyager</td>
<td>Voyager</td>
<td>Millennium</td>
</tr>
<tr>
<td></td>
<td>(called CLIO) with potential to move to Alma in coordination with Cornell. The Columbia Law Library uses Millennium.</td>
<td>- tentative plans to migrate to Alma but not until the new software is seen to be ready for general release (2013?). The library has not yet contracted to purchase Alma.</td>
<td>- possible interest in moving to Sierra at some point in future.</td>
</tr>
<tr>
<td>Discovery Tools (Note that all three discovery environment are based on SOLR.)</td>
<td>WebVoyage; CLIO Blacklight; and (Summon for e-resources but not for ReCAP resources)</td>
<td>WebVoyage - online catalog; Primo - for Voyager content; and (Summon - for e-journal articles but not for ReCAP resources).</td>
<td>WebPAC PRO and Bibliocommons</td>
</tr>
<tr>
<td>E-Document Delivery</td>
<td>RapidILL for article sharing (although ReCAP items are not eligible for RapidILL). The Avery Architectural Library is planning to use ILLiad for EDD.</td>
<td>RapidILL for article sharing. ILLiad for other EDD (called &quot;Article Express&quot;).</td>
<td>Ariel for ReCAP materials; email attachment for TSD.</td>
</tr>
<tr>
<td>Interlibrary Loan</td>
<td>ILLiad; and Relais for Borrow Direct.</td>
<td>ILLiad; and Relais for Borrow Direct</td>
<td>ILLiad</td>
</tr>
<tr>
<td>Link Resolver</td>
<td>360 Link</td>
<td>SFX</td>
<td>360 Link</td>
</tr>
<tr>
<td>Authentication</td>
<td>Columbia University maintains a central authentication service (PAM) which the libraries use in conjunction with EZ Proxy.</td>
<td>Voyager login credentials (barcode number and pin); IP range; VPN access; or proxy service.</td>
<td>Millennium and/or Bibliocommons login credentials: Barcode and PIN. No Shibboleth.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consortial/cooperative affiliations</th>
<th>Columbia</th>
<th>Princeton</th>
<th>NYPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Borrow Direct (Brown, Columbia, Cornell, Dartmouth, Univ of Penn, Princeton, and Yale)</td>
<td>☑️</td>
<td>❌</td>
<td>🌶</td>
</tr>
<tr>
<td>Rapid ILL</td>
<td>☑️</td>
<td>❌</td>
<td>🌶</td>
</tr>
<tr>
<td>MaRU (Manhattan Research Library Initiative) NYU is the third partner.</td>
<td>☑️</td>
<td></td>
<td>🌶</td>
</tr>
<tr>
<td>IDS (Information Delivery Service in NY State)</td>
<td></td>
<td></td>
<td>☑️</td>
</tr>
<tr>
<td>2CUL (Columbia and Cornell)</td>
<td>☑️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Google Book mass digitization program</td>
<td>☑️</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Hathi Trust</td>
<td>☑️</td>
<td>❌</td>
<td>🌶</td>
</tr>
<tr>
<td>Workflow task</td>
<td>Columbia</td>
<td>Princeton</td>
<td>NYPL</td>
</tr>
<tr>
<td>---------------</td>
<td>----------</td>
<td>-----------</td>
<td>------</td>
</tr>
<tr>
<td>Library identifies ReCAP materials (note de-duplication policies)</td>
<td>Columbia uses system generated lists - based on pub year and circulation history (like: published prior to 2003 and no circ since 2003). Selectors review and &quot;rescue&quot; items they think should remain on campus. Projects are underway in Music, Business and Butler - where the 12th Tier is being cleared. No circulating duplicates are intentionally sent to ReCAP. They are withdrawn instead. For many librarians, a scanned ToC increases the acceptability of sending materials to ReCAP. Pre 1830 materials are always assigned to the Rare Books and Manuscripts customer code (RS) or similar restricted codes for other special collections.</td>
<td>A Voyager report of holdings (by title, arranged by call number) is output by Tech Services for review by selectors. The lists always concentrate on a particular subject area and limited by predetermined criteria - set in consultation with that subject's primary selector. Typical criteria might be &quot;acquired more than [10] years ago and titles which have not circulated since [2000].&quot; These criteria are influenced by the target number of volumes required for relocation - and for serials, whether or not an online surrogate is available for all or part of a title's volumes. Once the report is run, selectors interact with the list via a locally developed (Access based) review tool. Selectors choose &quot;ReCAP&quot;, &quot;Withdraw&quot; or &quot;Keep&quot; - with ReCAP being the default decision. Princeton's policy is that no circulating internal duplicates are sent to ReCAP.</td>
<td>There are some item specific decisions - but more typically, whole ranges/sections of the collection are consigned to offsite storage. Example: New non-English titles. Internal duplicates DO get sent to ReCAP (Brewster). The Library's Deaccessioning Procedures are quite complex but for practical purposes, nothing can be withdrawn unless it can be agreed that it has no useful life and no artifactual value. Note that the customer code QJ is currently being used for selected second and third copies of SIBL journals (materials that NYPL would like to de-duplicate eventually.)</td>
</tr>
<tr>
<td>ReCAP decisions communicated to stacks or project managers</td>
<td>Smart bar-codes are generated for material consigned to ReCAP. The ReCAP customer code is the barcode prefix and a color strip also correlates with the customer code. The biggest customer code problem at present is CR insofar as it comprises too many different kinds of restrictions.</td>
<td>Picklists are generated from the Access file - one for transfers and one for withdrawals. Note that branch libraries work without benefit of this system - but manually apply whatever criteria they determine to be appropriate.</td>
<td>A decision must be made about which customer code will be assigned to each part of the collection - Collection Strategy makes these decisions. Some legacy decisions are in place for ongoing transfer to ReCAP of some classes.</td>
</tr>
<tr>
<td>ReCAP materials picked and verified</td>
<td>Some materials have an existing Item record with a local barcode. If an Item record doesn't exist during smart barcoding, one is automatically created. If extent the smart barcode is automatically activated. Processing sets and serials requires dumb barcodes assigned volume by volume. Item records are created anew and old ones deleted as appropriate.</td>
<td>Items are loaded onto books carts for shipment to ReCAP on a daily basis. These materials continue to have an &quot;available&quot; status throughout the transport and accessioning process.</td>
<td>A bar code for each item is scanned and an item/bib record match is confirmed. Note that only about 50% of NYPL holdings have barcodes. There are additional steps for items without. NYPL has only been barcoding for about 10 years.</td>
</tr>
<tr>
<td>Record maintenance</td>
<td>Staff verifies a bibliographic match and completes the record structure (holdings and items) as necessary. For small scale recon projects, the CLIO location is changed to a transitional location (like gla4off).</td>
<td>ReCAP processing of older holdings usually requires item level catalog maintenance and barcoding. Princeton has a Voyager location code for each ReCAP customer code - which allows them to display accurate borrowing rules in the catalog.</td>
<td>Each item is checked out to a pseudo patron.</td>
</tr>
<tr>
<td>Workflow task</td>
<td>Columbia</td>
<td>Princeton</td>
<td>NYPL</td>
</tr>
<tr>
<td>---------------</td>
<td>----------</td>
<td>-----------</td>
<td>------</td>
</tr>
<tr>
<td>ReCAP materials marked and prepared for ReCAP</td>
<td>An offsite (smart) barcode is applied to the top left corner of front cover. These barcodes have the appropriate ReCAP customer code as the prefix. For new acquisitions, the verso of the title page is marked with &quot;Offsite&quot; above the call number. Tyvek envelopes are used to house fragile material. These envelopes are printed with the crown logo and &quot;Columbia University Libraries Off-Site&quot;. They close with button and string fasteners. Acid free cardboard is added for flimsy objects. Cotton tape/string is used to secure damaged volumes with separate hinges or badly damaged bindings. Fragile books are staged and prepared separately. Smart barcodes are replaced with dumb barcodes for these materials. Transportation usually occurs shortly after preparation. If delays are expected, prepared books may be staged for retrieval by staff in response to patron requests.</td>
<td>The item's local barcode is already positioned on the top left corner of the front cover, which matches the ReCAP requirement. A pink ReCAP sticker is placed on the spine. Mylar bags are used to house fragile and multipart items - through which the items can be seen and the barcodes can be read. This obviates the need for barcoding the housing separately from the object itself. Princeton no longer writes the call number in the book.</td>
<td>For items with existing barcodes, a duplicate barcode is created and applied to the top left corner of the front cover. For those without, a new one is applied there and scanned into the item record. ReCAP call numbers (that look like this: 12-xxxx) are written inside the book and printed on spine labels. Pamphlets (&lt;50 pages) are put into acid free envelopes. Fragile materials are housed in Tyvek envelopes, tied with acid free boards and archival tying tape; or housed in custom enclosures. Eight different sizes of custom Tyvek envelopes are imprinted with the NYPL logo.</td>
</tr>
<tr>
<td>Materials transported to ReCAP</td>
<td>Materials bound for ReCAP are transported by Clancy-Cullen. Books are loaded onto wooden transfer carts - each one labeled with a blue tag indicating the ReCAP customer code and the delivery date. Truck loads are sent to ReCAP 3-5 times per month. (Approximately 4,500 items per truck load.) New acquisitions processed for ReCAP are sent daily via the Bohrens courier.</td>
<td>Princeton's ReCAP quota is 12 carts per day (carts hold approximately 125 volumes). Carts are labeled with the appropriate ReCAP customer code and delivery date. These are transported to ReCAP by Princeton Shipping once a day at around 2 pm. Shipping makes a daily run around the library system picking up new acquisitions from the branches - some of whom only send a load once or twice a week as they do not always process enough items to fill a cart each day. No ReCAP requests are picked up on this run.</td>
<td>At present, most of NYPL's processed materials are going to the Brewster facility - to be transferred at a later date to ReCAP. As many as possible are accessioned under the customer code XA. Approximately 65,000 items per week are transported (by Clancy-Cullen.) NYPL has replicated ReCAP customer codes except when NYPL/ReCAP customer codes are already assigned in the Brewster GFA system - notably: NA. Additional customer codes have been needed for billing purposes (QA, QJ for SIBL). QAs will likely need to be re-trayed and accessioned into NA when relocated to ReCAP. When XAs are transferred to ReCAP, whole trays should be able to be accessioned at ReCAP (GFA has precedent for this). XA is not currently assigned at ReCAP and should be reserved for eventual NYPL use.</td>
</tr>
<tr>
<td>Rejects returned to the libraries approximately once a week. These may be too badly damaged, missing barcodes, dirty, moldy, etc.</td>
<td>There is no pre-set schedule for rejects - are simply sent whenever they are encountered by ReCAP. Approximately 400 during 2012.</td>
<td>Approximately 250 in 2012.</td>
<td>&quot;Barcode not in ILS&quot; items are requested daily from TSD for correction. TSD sends missing or duplicate barcodes or other rejects weekly. Approximately 100 problems surface each week.</td>
</tr>
<tr>
<td>Materials sized by ReCAP - and located in trays.</td>
<td>Because these deliveries arrive in large volume (by the truckload) there can be up to a week of lag time before materials are sized.</td>
<td>Princeton materials are sized daily with no lag time.</td>
<td>Again, this is now happening for the most part at Brewster. At present there is no backlog. Brewster provides same/next day accessioning for 12,000 - 13,000 items daily.</td>
</tr>
<tr>
<td>Workflow task</td>
<td>Columbia</td>
<td>Princeton</td>
<td>NYPL</td>
</tr>
<tr>
<td>---------------</td>
<td>----------</td>
<td>-----------</td>
<td>------</td>
</tr>
<tr>
<td>Materials are accessioned by ReCAP. This involves scanning the barcode for each (previously sized) tray and for each item in the tray. The system assigns an aisle and shelf location. A work order number is assigned to each batch and item counts are recorded on work order slips. This is currently happening at a rate of 3,000 items per day. At peak load, the facility was accessioning approximately 8,000 items per day (with double the current staff). The full capacity of this function depends on the projected accessions - ReCAP hires accordingly.</td>
<td>Columbia staff expect a 2-4 week timeline between physical transfer and requestability by patrons in the OPAC.</td>
<td></td>
<td>Accessioning by tray could significantly increase ReCAP’s ability to ingest NYPL materials currently in the Brewster facility (see note above).</td>
</tr>
<tr>
<td>Accessions are &quot;verified&quot; - which basically means that the accession step is repeated and the work order is 'closed' before shelving.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reports of newly accessioned items (barcodes) are delivered via ftp to libraries daily.</td>
<td>Newly accessioned barcodes are retrieved weekly (on Fridays).</td>
<td>Nothing is currently done with these lists at Princeton.</td>
<td>Daily list of barcodes received from Brewster.</td>
</tr>
<tr>
<td>Record maintenance</td>
<td>As a batch process, the location code for accessioned items is flipped to an offsite location (off.glx). The second part of the code is the owning library. This causes the Request Button to appear in the OPAC and in Blacklight. Selected categories of the Accession report (&quot;to be discharged&quot;; &quot;temp location&quot;; etc.) are emailed to staff for adjustment. MRP staff resolve problems and complete CLIO updates.</td>
<td>Princeton has not implemented any routines to check the ReCAP accessions lists against Voyager.</td>
<td>The list of items accessioned at Brewster is used to drive a daily batch &quot;check-in&quot; process - which changes the status back to &quot;available&quot; and updates the Location, the OPAC message, and the Item Agency fields. The Location reflects offsite + originating library building/collection; Agency reflects customer code; and OPAC Message = &quot;ADV REQUEST&quot;.</td>
</tr>
</tbody>
</table>
### ReCAP: Current Workflows (July 2012)

<table>
<thead>
<tr>
<th>Workflow task</th>
<th>Columbia</th>
<th>Princeton</th>
<th>NYPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ReCAP materials discovered</td>
<td>ReCAP materials can be discovered in the OPAC or in Blacklight.</td>
<td>ReCAP materials can be discovered in either WebVoyage or Primo.</td>
<td>At present, there are two different discovery environments - Bibliocommons is the default (and most heavily used) view, but the classic catalog interface (WebPac PRO) is also available. As of May 2012, ReCAP materials are displayed in the WebPac as “requestable”. Note that the ReCAP customer code does not reside in Millennium, so it is not currently possible to inform the patron correctly about non-standard access options. If we understand correctly, the ILS office is doing an audit that will update all items with an Agency code reflecting the Customer Code under which the barcode was accessioned. This process will also update the location code to reflect any delivery restrictions, allowing the patron request form to display the allowed pickup locations.</td>
</tr>
<tr>
<td>ReCAP materials requested</td>
<td>An online request form form ReCAP materials has been integrated in both Blacklight and the OPAC. Columbia maintains a local indicator of status for items that have been requested from ReCAP, but have not yet been retrieved. This is understood to be a cumbersome work-around, because there is currently no way to systematically retrieve LAS statuses in real-time.</td>
<td>Princeton has integrated an online form in both environments. (The underlying script for the form was developed at Columbia.) This form is displayed to patrons for materials coded with ReCAP locations. The form collects patron and item information and lists the valid pick-up locations. The form does not make direct updates to the Voyager item status or patron record.</td>
<td>An online request form has been implemented in the WebPac environment. This is an unmediated service (referred to as ‘patron-initiated’) that will be extended to Bibliocommons within the next 6-8 months. At present, however, hand written request forms are still in use for the majority of requests. These are sent downstairs by conveyor and hand carried to room 50 where they are keyed into the online request form by SASB supervisors. Also note that the ReCAP request process is just one of several retrieval workflows including onsite materials, offsite materials, call-ahead service, patron-initiated requests, and staff-only requests.</td>
</tr>
<tr>
<td>Requests are transmitted to ReCAP in comma delimited files, with quotes surrounding each data element. Required data elements include: a request code, the item barcode, the pick-up location, the default pick-up location, and the type of delivery.) Many other data elements are optional.</td>
<td>Requests are aggregated into a text file (in a format specified by ReCAP) that is transmitted to ReCAP 3 times each day, at 7:15 am, 11:45 am, and 2:45 pm. The data sent relates to the item (or article) being requested as well as the person placing the request.</td>
<td>The PUL System empties the ‘request basket’ and transmits contents to the ReCAP server 3 times per day, at 7:15 am, 11:45 am, and 2:45 pm. Note that the request process does not automatically update Voyager. We assume that the requests transmitted to ReCAP must include patron ID - since there is enough information printed on the ReCAP slip to place the item on hold. (See below.)</td>
<td>A file of requests are sent to ReCAP 3 times per day, at 8 am, 12 pm and 2:30 pm. (TSD requests are transmitted at 1 and 5 pm.) Requests are “rejected” if the item is already checked out (just a handful a day), but the patron is “blind” to the rejection (The fact of a reject is communicated to the patron on the next day - see below.) There are typically a couple hundred request per day.</td>
</tr>
<tr>
<td>Workflow task</td>
<td>Columbia</td>
<td>Princeton</td>
<td>NYPL</td>
</tr>
<tr>
<td>---------------</td>
<td>----------</td>
<td>-----------</td>
<td>------</td>
</tr>
<tr>
<td>ReCAP uploads requests. The GFA system is updated - requested items are set to 'RET' for retrieval. Rejected items (errors) are removed from the list and are reported to each library at the end of the day along with other status updates. This request activity report includes successful requests, rejects, and EDD links. This report is sent daily to the owning institution and includes any request of its items, including those made by other institutions or through ILL.</td>
<td>Error reports are processed to notify the requesting patron that a request is not being filled. Other statuses are not acted upon. Columbia weeds the local tracking files daily based on current GFA item status. Columbia staff note that very few requests fail. Usually failures are for EDD and relate to poor physical condition or bad citation.</td>
<td>Error reports are reviewed by library staff. Problem categories include: requested items that are not actually available to the specific patron; requested pick up location is invalid (for items from restricted locations (eg. Rare Books) that must be delivered to those original locations); item number does not exist at ReCAP; or the item has already been requested or loaned. Princeton receives 5-6 errors each day - and nearly all can eventually be found in-house.</td>
<td>When NYPL receives the daily request activity reports from ReCAP and TSD, the following automated message is sent to the patron's email address: &quot;You must contact your library delivery location to confirm delivery and to set up an appointment to view your requested material/s. Failure to schedule an appointment before your arrival may result in your material being unavailable due to delivery processing and librarian availability.&quot; (Is this still true?) If a request is &quot;rejected&quot; by ReCAP or by Brewster, and an email address exists for the patron, a rejection notice is sent.</td>
</tr>
<tr>
<td>Requests are picked at a rate of 65/hour/staff and are processed 3 times per day - at 7:30 am, 12:00 pm, and 3:00 pm. Each period will have a different number of staff pulling requested items, depending on the volume of requests. All requested materials are retrieved during the next period. Retrieved books are sorted for either courier service or electronic document delivery.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Items are sorted and scanned into barcoded totes for shipment. A manifest is printed and included for all the items in the tote. All manifests (for all three libraries) include the same data. Totes are tracked in the system to help ReCAP know which totes are at each library.</td>
<td>CUL Shipping keeps a local file of the barcodes and destinations of all incoming totes. (Every ReCAP tote has its own barcode assigned, using a symbology completely different from the shared collections). Plastic totes, metal carts, wooden carts, &quot;suitcases&quot; and bubble wrap are used for various types and sizes of collection.</td>
<td>An &quot;En Route&quot; delivery notice is generated by ReCAP and delivered to the patron. Currently, this automated email reads &quot;In Process&quot; which may be confusing to patrons. Perhaps the language should be changed to &quot;En Route.&quot;</td>
<td></td>
</tr>
</tbody>
</table>
### ReCAP: Current Workflows (July 2012)

<table>
<thead>
<tr>
<th>Workflow task</th>
<th>Columbia</th>
<th>Princeton</th>
<th>NYPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ReCAP requests are transported to libraries by courier either once or twice a day.</td>
<td>Bohrens Moving and Storage provides a 4 am transport service Monday through Friday. Tuesday is the highest-volume delivery day - as all the Friday afternoon and weekend requests are delivered on Tuesday. ReCAP requests are delivered to CUL before opening, except for university holidays and weather emergencies. (ReCAP returns are picked-up and delivered back to ReCAP during this same transport cycle.)</td>
<td>Princeton Shipping picks up requests from ReCAP once every week day at 12:30 pm. These are delivered around the branch circuit, theoretically delivering all requests by 2:00 pm, ending with Firestone.</td>
<td>Bohrens Moving and Storage provides a 4 am ReCAP transport service Monday through Friday and Clancy-Cullen performs the Saturday morning transport. ReCAP requests are dropped off at the library Tuesday through Saturday and delivered to four research locations by 9 am. Note that the Saturday morning delivery of Friday’s ReCAP requests means that there is nothing to deliver on Monday morning. (ReCAP returns are picked-up and delivered back to ReCAP during this same transport cycle.) Clancy-Cullen performs this service for materials at the Brewster facility.</td>
</tr>
<tr>
<td>Requested items are received by libraries</td>
<td>There are various circulation desks throughout the Columbia library system. Many of them process ReCAP materials. After verifying contents of the tote with the printed manifest, a pink ‘ReCAP’ sticker is applied to the spine.</td>
<td>When totes are received from ReCAP, circulation desk staff manually place the items on hold for the patron - based on information on the printed ReCAP slip. Is there a way for the ILS “hold” to be placed automatically?</td>
<td>Deliveries for GRD/Main Reading Room go to Room S0 for processing. Totes delivered to other stops are processed by those departments. Items arrive in plastic totes with a packing slip in each, which includes item barcode, patron name, and item call number. Items that arrive from TSD will arrive with pink TSD label, to distinguish them from ReCAP materials that (for the most part) have no label. A yellow ReCAP label is then applied to ReCAP items. Items are matched against the packing slip; staff check in each item; the status of each item is manually updated to ‘on holdshelf’; and a date-stamped slip - with the patron’s last name is inserted in each item. Placing the item on ‘hold’ changes the status to ‘not available’. These items are then shelved on the Hold Shelf alphabetically by patron name. The Hold Shelf is in the RMRR Enclosure (service desk).</td>
</tr>
<tr>
<td>Patron notification</td>
<td>An email notification is sent to the patron (using rus*), and a patron/date slip is inserted in the piece. The email notification explains that the item will be held for two weeks. These items are then charged to the ReCAP Hold Shelf status patron. Items are shelved alphabetically by patron name. The hold shelf is weeded daily - for items that have been there for two weeks.</td>
<td>Once the item is placed on hold for the patron, an automatically generated “item available” notice is sent to the patron. Notices are generated twice daily - once in the morning and once at 4:30 or 5:00 pm. Again, the goal is that all email notices have a time-stamp no later than 5:00 pm on the first business day after the request was placed.</td>
<td>If there is an email address on file for the requesting patron, a detailed email notification is sent that looks like this: “The item(s) you’ve requested is available and has been reserved for your use on [date]. This item will be held for you at the Delivery Desk in the Rose Main Reading Room (Room 315) of the Stephen A. Schwarzman Building on 5th Ave. &amp; 42nd Street until [date]. Please bring this confirmation at the time of pick-up.” (At present, these are created manually.) If there is a phone number and no email, a phone call is placed with a similar message.</td>
</tr>
<tr>
<td>Workflow task</td>
<td>Columbia</td>
<td>Princeton</td>
<td>NYPL</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Patron pick-up</td>
<td>Any patron with a UNI can place a physical delivery request. Patrons with borrowing privileges may charge circulating ReCAP collections. Staff charge within a happening location in Voyager, typically Butler Circulation.</td>
<td>ReCAP items are charged via normal circulation processes.</td>
<td>Items are checked out to the patron’s library card. With the exception of MaRLI items borrowed by MaRLI patrons, these materials are for in-house use. Items must be returned and checked in by the end of the business day, using Millennium check-in. If materials are needed for another day, the status is changed back to ‘on holdshelf’ and reshelved on the Research Reserve Shelf in the RMRR Enclosure. (Other divisions and buildings may not use the same procedure.)</td>
</tr>
<tr>
<td>Patron returns the ReCAP item to the library</td>
<td>Items are discharged ONLY at the Butler Circulation Desk or the owning library for restricted locations (Avery, some Music locations, etc.). Staff verify that there is a pink ReCAP sticker on every return. Any holds/recalls are immediately routed to the destination circ desk. Notifications are typically sent to patrons on the day following the original discharge. Books are packed with green foam to prevent shifting, and tote lids are secured with a cable tie. The blue destination card is flipped to display ‘ReCAP’.</td>
<td>Patrons return ReCAP items to the Circulation Desk - which are recognized by their pink ReCAP labels. Items are then discharged. The ‘discharge’ must occur under a ReCAP login. Items are not requestable again until they are “refilled” at ReCAP.</td>
<td>ReCAP and Brewster returns are distinguishable via their yellow and pink labels. The item status is set to ‘in transit’ at the time of the return but the check-in is accomplished in batch the next day - prompted by the re-filing report from ReCAP/Brewster.</td>
</tr>
<tr>
<td>ReCAP items are returned to ReCAP in gray ReCAP totes.</td>
<td>Returns are picked up daily from the Shipping Room by Bohren’s Moving and Storage - Monday through Friday at around 4 am.</td>
<td>Princeton Shipping picks up ReCAP returns once every week day at approximately 10:30 am. No pick-up of ReCAP requests is made on this run.</td>
<td>Staff in RM 50 pack ReCAP and Brewster returns in their respective totes. Totes await pick-up around 4 am the next morning on the loading dock.</td>
</tr>
<tr>
<td>ReCAP items are re-filed into ReCAP modules- to their exact tray and shelf location. Confirmation of re-filing is sent to libraries nightly. Re-filing can be accomplished at a rate of 50/hour/staff member.</td>
<td>LITO staff run a weekly reconciliation to clear the item status of returned ReCAP items. All &quot;charged&quot; and &quot;in transit&quot; statuses are corrected in CLIO once ReCAP confirms that the items have been re-filed.</td>
<td>Nothing is done with the report of re-filing confirmation generated at ReCAP from GFA (LAS).</td>
<td>NYPL receives daily via FTP, a list of all barcodes re-filed the previous day at ReCAP (and Brewster). ILS staff creates list(s) of re-filed barcodes and performs a batch check-in during which the item status is flipped back to “available”.</td>
</tr>
<tr>
<td>A significant number of volumes may be backlogged in the re-file queue. However, appropriate status updates have been sent to libraries and recent modifications to the GFA software have been made to support retrieval from the queue. Books might remain in this backlog up to a week, but again, their status has been set to ‘IN’ and they are available for re-request on the day following their arrival. Every week, there are a few retrieval requests for these items.</td>
<td>Not everyone knows about the LAS system upgrade.</td>
<td>Not everyone knows about the LAS system upgrade.</td>
<td>There is no apparent re-file backlog at TSD.</td>
</tr>
<tr>
<td>Workflow task</td>
<td>Columbia</td>
<td>Princeton</td>
<td>NYPL</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Newly acquired materials to ReCAP</td>
<td>Selected by bibliographers. Approximately 30,000 new acquisitions per year are sent directly to ReCAP. Can we eliminate the (local) campus barcode when items are sent directly to ReCAP?</td>
<td>Approximately 15,000 items per year based on title-by-title selector decisions. The ReCAP location is established/assigned at point of order or during post-receipt approval review.</td>
<td>“Shelf-ready” foreign language monographs go offsite after being received and 'processed' at the Library Services Center in Queens. Any chance that these could go directly to ReCAP in future? (50 -100,000 items per year.)</td>
</tr>
<tr>
<td>Workflow task</td>
<td>Columbia</td>
<td>Princeton</td>
<td>NYPL</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>ILL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In 2007, ReCAP started receiving ILL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>requests. Each library forwards ILL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>requests to ReCAP where staff act as</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>members of each library's ILL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>department - with access to their</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILLiad queue for ReCAP material. ReCAP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>staff identify and retrieve materials</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and send them to the requesting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>library. Approximately 11% of all</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ReCAP usage is for ILL.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILL, in contrast to ReCAP workflows,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>is a centralized service at Columbia.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Users don't know and shouldn't have</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to know the difference between ILL,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borrow Direct, ReCAP, etc.&quot; How to</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>simplify the request process for</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>patrons?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iliad requests are transferred to the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ReCAP queue and for the most part,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>items are returned directly to ReCAP.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The ReCAP sticker helps to identify</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>these so that they can be visually</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sorted. There is a weakness in the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relais relay of information to GFA.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A Z39.50 request is transmitted,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>which includes the item ID but not the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>barcode on the printing slip. (ReCAP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>is set up as a printing location.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The barcode must be attached manually,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>or sent in a separate email. Any way</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to remedy this? (Princeton just</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>switched from URSA to Relais last</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year. URSA used to include the barcode</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in the request transmission.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication with patrons about</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unfilled requests could be more</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>streamlined.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILL requests from ReCAP are processed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>at ReCAP. ILL requests from TSD are</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shipped to the ILL department for</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>processing.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Electronic Document Delivery</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When book chapters or journal articles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>are requested, ReCAP staff pull the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>items, scan the requested content and</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>deliver it directly to the requestor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>via Ariel. These scans are made</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>available for 2 weeks and are not</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>retained.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>There is a weakness in the Relais</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>relay. A Z39.50 request is transmitted,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>which includes the item ID but not the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>barcode on the printing slip. (ReCAP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>is set up as a printing location.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The barcode must be attached manually,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>or sent in a separate email. Any way</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to remedy this?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If the ReCAP or TSD request resulted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in an electronic document scan, the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>daily request activity report</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>indicates such. On the same day, the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDD’s are created and links are</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>emailed directly to the patron.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fair use guidelines at ReCAP are not</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>entirely consistent with fair use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>guidelines at the campuses of the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>partner institutions. The partners</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>will need to agree on guidelines for</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fair use that will satisfy the patrons</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>needs as well as the desire of the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>institutions’ offices of general</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>counsel.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Will de-accession upon request from</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>selectors and for correcting errors in</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>customer code accessioning, etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>There is no systematic de-accessioning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>workflow in place at present.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A workflow is in place for permanent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>withdrawal from ReCAP of single items.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This is not a scalable process.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Permanent deaccession from ReCAP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At present - no policies are in place</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>for doing this except in cases of</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>one-off accessioning errors. Selectors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>or administrators have requested</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de-accession in very low quantity.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A policy framework is being considered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>for high-use items.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Will de-accession upon request from</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>selectors and for correcting errors in</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>customer code accessioning, etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>There is no systematic de-accessioning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>workflow in place at present.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A workflow is in place for permanent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>withdrawal from ReCAP of single items.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This is not a scalable process.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*rus is a locally developed program</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>for notifying patrons that ReCAP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>requests have arrived and are available</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>for pick-up. It can be used to</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>retrieve information about the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>barcode’s requesting user ID and</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>delivery location. It does not</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>function as a tracking system. (In</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>theory this mean that the request</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>transmission to ReCAP does not need</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to include patron name or other patron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>information.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;GFA and the ReCAP shelving system is</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in general too rigid and mono-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>functional. For example, making</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>changes to an accessioned item is</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time-consuming and costly. For</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>individual items, the only way to</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>change a customer code is to withdraw</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and re-accession it. Changes to</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blocks of material, from one customer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>code to another, is very expensive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>especially when GFA has to be</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>involved.&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The workflow described here is for</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the most routine process. There are</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>some variations for RMRR, staff-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>only requests, when items are needed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>for more than one day, etc. The process</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>may vary in other departments, other</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>buildings.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix I. ReCAP Project Technology Report
ReCAP Project Technology Report  
Research Collections and Preservation Consortium  

By Marshall Breeding, ReCAP Technology Consultant  

Revised – November 2012  

As part of a project to expand the vision of the ReCAP facility from a shared storage facility to a shared collection with enhanced access to the patrons of each of the participating libraries, an enhanced technology infrastructure needs to be implemented. An enhanced technical environment to support the next phase of ReCAP might include capabilities such as:

- Inclusion of materials available to end users at any of the three participating institutions in the shared ReCAP collection through existing online catalog or discovery interfaces  
- Display of real-time status of items in ReCAP, including availability for request, restrictions, available pick-up locations  
- Improved services that can be embedded into the online catalog or discovery services of the participating institutions to capture and validate requests made by library users or by library personnel for ReCAP materials  
- Tools for tracking ReCAP materials requested by patrons from the time that they leave the ReCAP until they are returned for refilling  
- Tools to support the management of the ReCAP collection, such as collaborative collection development, automated processing of duplicates, or designation of preservation retention.

This report provides an environmental scan of the relevant technology components in place in the participating institutions and posits functionality to support the operation of ReCAP as a shared collection with unmediated patron request capabilities. The report suggests some general approaches and specific products or projects that might be considered as candidates as part of an expanded technical infrastructure.

Part I. Strategic Infrastructure used by ReCAP Institutions  

ReCAP Facility  
The key technology issue for the ReCAP facility itself centers on the Library Archival System developed by Generation Fifth Applications that provides a comprehensive inventory management system and that automates all of workflows carried out in accessioning new materials, handing retrieval requests, reporting statistics, and a variety of other tasks involved in the operation of the facility.
Library Archival System
The Library Archival System was developed by Generation Fifth Applications as an inventory management system based on the Progress fourth-generation language. This software tends to be called “GFA” by those using it.

Information regarding the GFA software was obtained through interviews with Eileen Henthorne (Executive Director ReCAP), Steve Gilbert (GFA President and owner) and Christine Brennan (GFA Operations Manager), as well as through other ReCAP participants.

Generation Fifth Applications (http://www.gfatech.com/) has been in business for about 30 years and develops applications for other industries besides that of library storage facilities.

Gilbert reports that the company considers itself very progressive in its approach to technology and is interested in new developments that would meet the needs of its library customers. He states that access to the database management system can be accomplished in any way that the libraries need. In addition to the batch mode that ReCAP uses for most of its work, direct mode is also available that provides more dynamic access to the data.

GFA Library Archival System is currently based on Version 10 of Progress (www.progress.com), with development underway for version 11. Most customer sites have migrated from Version 9 to 10.

Gilbert considers GFA as a research and development company. He reports that the company works in other areas, such as in telecommunications. The company owns multiple data centers. GFA was an early proponent of cloud computing concepts.

GFA follows a shared source business model. While customers do not receive the source code to the system, new functionality sponsored by one customer is shared with any other sites interested in the same functionality.

The development approach is to avoid being driven by the technology but by the operational needs of the organization.

Work is underway to expand the sub-catalog, developing a catalog in its own right. Such a catalog would be limited to inventory data and not bibliographic descriptions.

Gilbert indicates that the company would be generally open to creating additional functionality, such as service layer of APIs for ReCAP.

Functionality
The Library Archival System is designed to be an industrial inventory management system with operational requirements that differ considerably from an integrated library system. Its primary function is to automate the operations of the facility and to manage its inventory in such a way that no materials are lost.

The GFA software uses a text-based terminal interface accessed through a secure shell client (see www.ssh.com). The use of ssh uses an encrypted communications to provide secure access to the
software without exposing usernames, passwords, or operational data. All data entry is accomplished through menus and keyboard commands.

The Library Archival System and the operational procedures in place in ReCAP are designed to result in a very high level of control so that no materials are lost within the facility. ReCAP personnel have a very high degree of confidence in the system to manage their inventory and to automate their operations.

The boundaries of the control of ReCAP and the Library Archival System are rigidly defined. It controls items once they are received into the facility and accessioned into the GFA software. It exerts no control of items once they leave the facility. Libraries are responsible for tracking requested items from the point where they leave ReCAP until they are returned for re-filing. The handling of materials outside of the control of GFA leaves considerable room for uncertainty of status and availability and error.

**Data Structures**

Some of the data elements include a unique identifier, customer codes, location, current status information, and its transaction history.

Customer Codes are assigned to each accessioned item. These codes reflect the original institution and the collection or sub-collection. The customer code controls the business rules that apply to the item. Some customer codes represent materials that have special restrictions and may not be able to be requested by other institutions or even by some categories of users at the home institutions. Of special interest are the customer codes that can be requested by all the ReCAP participants that would comprise the shared collection.
The GFA Library Archival System maintains a detailed history on each accessioned item. This transaction log can be used for problem solving and for reports involving the frequency of use.

Within the ReCAP facility and within LAS, there is no representation of bibliographic data for any item. Bibliographic data remains in the respective integrated library systems.

**System Architecture**
The software has been created with the OpenEdge Advanced Business Language, commonly called Progress, a fourth-generation language that includes integrated relational database functions.

The Library Archival System, as are all products created by GFA are based on a framework called ToolPro, positioned as a fifth-generation language. The name of the company derives from its orientation toward this fifth generation language for software development.

GFA describes the LAS application as comprised of at least a hundred smaller programs, which have been assembled to meet the requirements of ReCAP. Each implementation of LAS will be different, depending on the specific programs needed for their operations and workflows.

ReCAP, for example uses a program that performs “transit bin tracking” which is used only by about a third of organizations using LAS.

LAS can be delivered to support different workflow models. The software has been implemented by the University of Oxford in the United Kingdom, for example, and their version has been configured to handle a different workflow for the accessioning of materials. Oxford uses a linear accessioning process to accomplish an accession rate of 40,000 items per day to meet their goal of processing 7 million items per year. This linear workflow involves having materials pre-sized so that a group of three personnel work together on a workbench to do accessioning and verification. In contract, ReCAP uses a parallel workflow involving three sets of operators working separately.

The GFA software is not perceived to be very open by the libraries involved in ReCAP. While there are many reports available and the ability to perform real-time queries, there are no APIs currently available to customers to extend functionality or for interoperability scenarios.

**Deployment model**
The GFA Library Archival System is operated on a virtual server housed at Princeton University.

No problems have been reported with this arrangement. The implementation has proven to be extremely reliable with very little down time.

The implementation can be considered a server-based system operated through character terminal interfaces accessed through secure shell clients.

**Enhancement Process**
While incremental enhancements are done as needed, no plans are currently underway for a major revision of the software or to consider other alternative systems.
The last major revision to the software was performed in around 2001. Since that time incremental revisions have continued to be implemented upon request from the broader GFA customer base.

Brennan indicates that there has been some interest in providing a Web-based or graphical interface for LAS. This may not be needed for operational staff of the facilities that require very efficient keyboard workflows. Use of a mouse would likely reduce efficiency. Other types of use, such as those by authorized staff from the participating libraries, or those that need to create reports would benefit from a friendlier interface.

Other future developments might include the ability to operate the software from iPads or other tablets. This capability would provide support for lower-cost devices for real-time access from within the storage modules, when wireless network access is already available.

Eileen Henthorne indicates that some of the desired enhancements that would benefit ReCAP operations might include the ability to do SQL-like queries and to have more capabilities for ad-hoc reports. But from the perspective of operational efficiency, the software in its current state is believed to work quite well.

**Customer base**

GFA Library Archival System finds use in most of the major high-density, non-robotic library storage facilities. Major installations include the Harvard University Depository, Cornell University Library Annex, and many others are listed on the GFA Web site. The system has recently been implemented to support Oxford University Bodleian Library’s Book Storage Facility in Swindon, representing the largest installation to date. The Clancy-Cullen facility in upstate New York, with which NYPL contracts for the short-term transfer and storage of their materials, also uses the GFA system.

Large-scale library repositories represent a very small niche for automation support. While these facilities may have some characteristics in common with other types of warehouse operations, they have very specialized needs. There are not currently other products that directly compete with LCA. The only competition is a conceptual one where some libraries have chosen to use their integrated library system to manage the inventory of their storage facility.

There is not a formal user group organized for the institutions that have implemented the GFA Library Archival System, but there have been some informal meetings.

**Cost and Licensing**

The Library Archival System is licensed according to the number of staff clients, or “seats” that can access the system simultaneously. The initial cost of the software was around $100,000 [need specific figures] and the annual maintenance is approximately $16,000. The maintenance includes problem support as well as access to new versions and enhancements.

**Observations about Library Archival System**

The number of large-scale repositories using the GFA software reflects its dominant position. It does not seem likely that other applications will emerge from competitors given the number of potential clients would be limited relative to research and development costs. The GFA software has been developed
primarily for the operational needs of the repositories. Other concerns such as interoperability with discovery services and workflow management systems have not previously been part of the development strategy. The extension of the software for these purposes would not only benefit ReCAP, but would likely be of interest to other institutions.

ReCAP Functional Workflows
All tracking in the ReCAP facility is accomplished through barcodes. All materials submitted to ReCAP must have barcodes correctly positioned. The rapid processing off material depends on the consistent placement of barcodes. All containers involved in the process have unique barcode numbers such as bin, transportation totes. Each shelf in the storage modules has a barcode.

No alternative identification technologies, such as RFID, are used by ReCAP.

One of the key principles of high-density storage facilities involves making efficient use of storage space by organizing materials according to size. ReCAP stores items in trays according to factors such as size and the owning institutions.

No consideration is given to bibliographic details or classification schemes. One of the key issues for ReCAP and the GFA software is that associations between inventory data and bibliographic data must be handled through a hybrid system involving the ILS of each library.
The accessioning process begins with sizing items and assigning them to size-related categories. A given tray is filled with materials of its designated size classification. Each tray is barcoded with a unique identifier and each item placed in it is scanned and registered accordingly to that tray.

![Sizing materials as part of accessioning process](image)

Figure 2 Sizing materials as part of accessioning process

The GFA system assigns a location for the bin based on its size and available space of that category. Configuration parameters and a data representing all available space in the facility drive the system to assign shelf locations.

All trays are verified by re-scanning to ensure that all items are present.
The accessioning process is designed to ensure that any given item can be definitively located, including the module, aisle, shelf location, and bin within the shelf.

The totes used to ship materials to the libraries are also barcoded and tracked with an optional module to help ReCAP know which totes are at a given library.

**Transmission of requests**

Requests from the libraries are currently managed through the exchange of data files in scheduled batch operations. Each library will create a file of requested items. The file is created through scripts at each institution that assemble the accumulated requests received during that period.

Files are structured in comma delimited format, with quotes surrounding each data element. The following data elements are required in files submitted to ReCAP:

<table>
<thead>
<tr>
<th>Field</th>
<th>Required?</th>
<th>Content*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Request code</td>
<td>y</td>
<td>REQI, ADDI, DELE</td>
</tr>
<tr>
<td>Item barcode</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>Priority code</td>
<td>n</td>
<td>I</td>
</tr>
<tr>
<td>Patron name</td>
<td>n</td>
<td></td>
</tr>
<tr>
<td>Item title</td>
<td>n</td>
<td></td>
</tr>
<tr>
<td>Pick-up location</td>
<td>y</td>
<td></td>
</tr>
</tbody>
</table>
ReCAP sends back to each library a file that reflects errors and the status of requests.

Each of the participating institutions performs updates to the status of items within its ILS according to these transactions. A successful request to ReCAP, for example, will cause the owing institution’s ILS to flag the item as unavailable. Status updates are sent to the owning institution for any request of its items, including those made by other institutions or through interlibrary loan.

Rather than gathering requests into a single daily batch file, other procedures have been tried in the past that included processing individual requests as they happen. This method proved to be problematic.

When libraries receive materials from ReCAP that have been requested by patrons, they use their local ILS to manage notification and check-out. A typical workflow would involve a staff member using the holds management feature of the ILS to associate the item with the patron record and to issue a notice that it is available. The item would be checked out to the patron through the circulation module.

This workflow may need to be adjusted when patrons have access to ReCAP materials from the other two libraries that do not already exist within their ILS. Some process of automatically moving bibliographic and item records for requested items into the local ILS for circulation would facilitate the fulfillment of these materials to patrons.

Another set of workflows applies to the return of requested materials from patrons, through their libraries and ultimately to ReCAP for refilling. The priority of refilling materials back into the storage modules is naturally not as high as satisfying requests. A significant number of materials may be backlogged awaiting refilling. Materials in this state receive appropriate status updates and modifications of the GFA software has been made to support retrieval, including the ability for ReCAP personnel to know the specific location of items among the refilling backlog.

<table>
<thead>
<tr>
<th>Default pick-up location</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item call number</td>
<td>n</td>
</tr>
<tr>
<td>Item author</td>
<td>n</td>
</tr>
<tr>
<td>Item vol/part/date</td>
<td>n</td>
</tr>
<tr>
<td>Patron email</td>
<td>n</td>
</tr>
<tr>
<td>Other Patron info</td>
<td>n</td>
</tr>
<tr>
<td>Article vol/part/date</td>
<td>n</td>
</tr>
<tr>
<td>Article author</td>
<td>n</td>
</tr>
<tr>
<td>Article title</td>
<td>n</td>
</tr>
<tr>
<td>Start page</td>
<td>n</td>
</tr>
<tr>
<td>End page</td>
<td>n</td>
</tr>
<tr>
<td>Other identifying info</td>
<td>n</td>
</tr>
<tr>
<td>Notes</td>
<td>n</td>
</tr>
<tr>
<td>Type of delivery</td>
<td>y</td>
</tr>
</tbody>
</table>

Table of data elements provided by Princeton University Library
Interlibrary Loan
In addition to the requests directly made from each library, ReCAP supports interlibrary loan requests. If an interlibrary loan request is made for an item held in ReCAP, an expedited process has been developed to allow it to be processed directly from the facility rather than having to be first retrieved by the library. ReCAP maintains an interlibrary-loan workstation for each of the institutions and its personnel fulfill these requests daily.

Electronic Document Delivery
A substantial number of requests can be satisfied through electronic document delivery. For book chapters or journal articles, ReCAP personnel pull the items, scan the article or book chapter, with direct delivery to the requestor.

To satisfy these requests, materials are retrieved from the storage modules and delivered to an operator that scans the requested pages. The electronic file is then submitted for delivery through the document delivery management application, using the Ariel document delivery software provided by Infotrieve.
Duplication issues
One operational issue for ReCAP involves duplication of materials. The ReCAP facility currently includes some duplication of items across the three institutions. There is some desire to reduce the amount of duplication, but this is more of a library issue, in that they pay according to the number of items accessioned than an operational issue for ReCAP.

Collection analysis is underway to help identify duplication of materials among the participating institutions. It may be helpful in the future to have enhanced discovery tools for library personnel that span the entire collections of the participating institutions to assist in developing collections without accidental duplication.

ReCAP Governance
The ReCAP facility is governed through a board comprised of the University Librarian and Chief Financial Officer from each of the three institutions that comprise the consortium. ReCAP is separately incorporated as a non-profit. The personnel of ReCAP are employees of Princeton University.

Eileen Henthorne serves as the ReCAP executive director and reports to the ReCAP Board of Directors.
The scans made for electronic document delivery are not retained due to copyright concerns. It might be worth investigating whether these scans could be transferred to HathiTrust. It seems that an automated process could be developed to perform OCR, associate metadata, and other tasks to turn them into reusable digital objects that could be used under the same constraints as other items in HathiTrust. The cumulative amount of material collected is substantial; it would be a positive contribution to find ways to gain additional value from them.

**Technology Issues**

One of the key problems in the new environment will involve maintaining the status of any given item stored at ReCAP. Each of the individual integrated library systems indicates when an item has been transferred to ReCAP and its availability status. When the library requests its own materials, it naturally has the data needed to update that status.

When another library requests an item, ReCAP also sends a message so that the owning library can update its status in the ILS.

The lack of a modern service layer of application programming interfaces (APIs) is a key issue for this project. The new system needs to be able to interrogate the GFA software in real time in various scenarios regarding its current status and availability.

Access to the LAS through only through the menu-based character online interface and through the batch processes based on exchange of text files constrains how the software would fit into a modern service-oriented environment. While the character interface provides efficiency for internal operations of facilities such as ReCAP, the availability of a set of APIs implemented through REST services would help integrate the system into a broader technology environment for the broader consortium.

**New York Public Library**

New York Public Library currently has 3.5 volumes stored in the ReCAP facility, with plans to transfer a significant number of additional materials as part of its Central Library Project. This renovation project not only involves the transfer of many volumes out of the building, but also has a very high level of public attention, increasing the importance of service improvements gained through the next phase of the ReCAP project.

Many differences obtain between NYPL and the other two ReCAP institutions. Different use patterns apply to its research collection with materials primarily used only in its reading rooms while Princeton and Columbia allow most materials to be checked out to their users. These differences in policies will need to be taken into consideration as workflows are developed and as technology components are constructed in support of these workflows.

NYPL along with Columbia University and New York University formed a new cooperative relationship, called the Manhattan Research Library Initiative, or MaRLI. Through this new partnership, NYPL users otherwise not affiliated with Columbia or NYU will be able to arrange borrowing privileges at those
institutions. Reciprocal arrangements would also allow PhD candidates and faculty members of Columbia and NYU access to NYPL’s research collection. MaRLI will also involve some degree of collaborative collection development among these three institutions. MaRLI has only an indirect on ReCAP.

NYPL also participates in the IDS Project, a resource-sharing cooperative among a number of institutions in the state of New York. See: idsproject.org.

All three ReCAP institutions are involved as partners in the Google Book mass digitization program. NYPL announced its participation in December 2004; Columbia joined in December 2007; Princeton joined in February 2007.

Likewise NYPL, Princeton University, and Columbia University all belong to HathiTrust, a collaborative digital repository, which includes the library scans from the Google Project in addition to materials from other digitization projects among the members.

Automation Environment

The New York Public Library currently has an automation environment based on the Millennium integrated library system from Innovative Interfaces. Millennium was placed in production use in 2009 following an implementation that for the first time combined the NYPL Branch Libraries, migrating from a Dynix Classic system, and the NYPL Research Libraries, migrating from an existing Millennium system. The contract with Innovative to implement Millennium was announced in March 2008.

There is not a single library system or automation environment that spans all the boroughs of New York City. Queens Public Library uses a VIRTUA system from VTLS; Brooklyn Public Library uses a separate Millennium implementation.

NYPL makes use of multiple modules of Millennium, including Cataloging, Circulation, Acquisitions, Circa (wireless circulation and inventory), the WebPAC PRO online catalog, and Ecommerce (electronic online collection of fines and fees). Encore was originally licensed from Innovative, but it has since been replaced by BiblioCommons.

NYPL includes a large number of neighborhood branch libraries in addition to the research library. The BiblioCommons discovery interface appeals more to the type of use associated with the branch libraries. BiblioCommons, even though it is currently the default search presented on NYPL.org will not be positioned as the interface for the research libraries. The “classic catalog” based on WebPAC Pro from Innovative Interfaces will be the main search interface designated for the users of the research libraries.

NYPL does not have specific plans in place to replace Millennium, though there is some interest in moving to Sierra. The migration to Sierra could come as early as late 2012 or at probably within 2013. (At the August 21, 2012 meeting, NYPL indicated that it is planning for a transition in late 2012 or early 2013; by the September 11 call the implementation had been delayed until about June 2013.)
Depending on the timeframe for the implementation of a new technical environment for ReCAP, support will need to be in place for Millennium even if Sierra or another system will eventually be implemented for NYPL.

**Millennium and Sierra**

Innovative Interfaces developed its Millennium ILS beginning about 1997 as the forward path for its INNOPAC ILS. Millennium was developed with a modular client/server architecture offering staff clients created in the Java Swing framework and a Web-based online catalog. INNOPAP relied on text-based telnet clients and included very mature functionality, especially for technical services operations. Millennium carried forward that functionality into a graphical environment.

Millennium offers a very rich set of functionality across the main areas of library operations, including cataloging, acquisitions, serials management, and circulation, implemented through staff clients specialized for each module.

The initial implementation of Millennium included the use of the Encore discovery service developed by Innovative Interfaces, which has subsequently been replaced by BiblioCommons.

An important characteristic of Millennium involves the high degree of control that Innovative exerts on the system. It is a proprietary system with only limited avenues for organizations implementing the system to gain programmatic access to its underlying data and functionality.

Innovative offers some packages of application programming interfaces that libraries can license:

1. Patron API – procedures related to authentication, to query the patron database for records, to query or update patron record data, and to set or clear fines.
2. Item Status API – procedures related to the query of items, updating status, and performing check-outs
3. My Millennium API
4. NCIP – NISO Circulation Interchange Protocol used primarily in resource sharing environments, but also designed for self-service and other applications requiring access to patron and item functions.

Sierra represents Innovative’s efforts to create a next-generation automation environment based on current technology while moving forward all existing functionality. The following excerpt from *Smart Libraries Newsletter* summarizes some of the relevant characteristics of Sierra:

The launch of a new system affords Innovative the opportunity to take advantage of technology components, architectures, and methods consistent with the times, and hopefully to anticipate what features will be popular and necessary in the future. Innovative has chosen the service-oriented architecture, open source database and indexing components, RESTful web services and APIs, and engagement with library developers as some of the key elements of its new technology strategy.
Following a service-oriented architecture, the new Sierra platform has been constructed in four layers. The foundation database layer will make use of PostgreSQL as the transactional database engine and Lucene for indexing to support search and retrieval operations. Use of these open source components stands in distinct contrast to Millennium, which uses proprietary database and indexing technologies developed by Innovative, or the commercial Oracle database. The database layer connects to the rest of the application through a data access object layer using a component called Hibernate which lends database independence and the ability to maintain persistent transactions through workflows executed at higher levels. Since the database layer also supports standard SQL, third party tools such as Crystal Reports can be used to create reports or other data extraction or manipulation operations. Use of the open source PostgreSQL also results in significant cost savings relative to commercial database engines.

A services layer implements the business logic that represents all of the functionality of the system, including the detailed tasks and workflows involved in ILS modules such as cataloging, circulation, acquisitions, a set of services for managing electronic resources, and another others for discovery and delivery of content. The services of this layer are exposed to higher-level applications through SOAP wrappers.

A set of new Sierra applications sits on top of the services layer. The heart of this layer will be a new Sierra App that implements the staff functionality of the system through a unified, non-modular approach. While Millennium offered specialized clients for each of its functional modules, Sierra delivers all functionality through a single application, avoiding the need to switch among modules depending on the task at hand. This application layer will also include components to deliver bundles of API's to support a variety of external interactions, all delivered through RESTful web services (Representational State Transfer). A top presentation layer will operate above the application layer, including the client to the Sierra application, web and mobile public interfaces. This presentation layer would also include third-party applications built on top of the published API's, interfaces to social media applications, or other end-user applications that might be created.

The new Sierra platform will be offered as software that can be installed locally in a library or consortium and will also be offered through software-as-a-service, hosted in a cloud infrastructure.


NYPL Public Interfaces
NYPL offers patron access to its collections through both the WebPAC Web-based online catalog module of Millennium and through BiblioCommons.

Classic Catalog: http://catalog.nypl.org/
BiblioCommons: http://nypl.bibliocommons.com/

**BiblioCommons**

NYPL implemented BiblioCommons as its primary discovery interface in June 2011. BiblioCommons became the default catalog interface for NYPL beginning about September 2011.

BiblioCommons follows a modern service-oriented architecture and has been deployed as a multi-tenant software-as-a-service environment. The service is hosted on BiblioCommons servers and carries their domain name in the URL (nypl.bibliocommons.com).

The following excerpt summarizes the architecture and technology of BiblioCommons:

Like the majority of existing discovery interface products, BiblioCommons operates separately from the underlying integrated library system. Like the generic discovery interface model, data are harvested from the ILS and used to populate a separate search and retrieval environment. Products like Primo, Encore and AquaBrowser each harvest the metadata from the ILS and other local collections into an instance of the software specific to the implementation of a library or consortium. BiblioCommons takes a fairly radical departure in that data from the ILS of each participating library loads into a centralized site. From the perspective of the patron, the library may scope the search to a given library or region, but the fundamental concept of BiblioCommons involves broadly shared data. In addition to harvesting basic bibliographic records, BiblioCommons harvests holdings and item-level data as well as authority records. Even though BiblioCommons relies on a shared bibliographic database, it preserves and indexes any locally created cataloging.

One of the key issues with discovery interfaces is the way that they overlap and interact with the underlying ILS. BiblioCommons shares the concept of harvesting and synchronizing data describing the collection from the ILS, but into a collective service rather than library-specific implementations.

Some of the discovery interfaces tap into the online catalog features of the ILS for item-specific displays and services, such as placing a hold. BiblioCommons completely replaces the online catalog of the ILS, managing all aspects of the way that the patron interacts with the collection. Its emphasis on the patron and social interactions require a much more sophisticated approach than a simple hand off to the patron services functionality built into the online catalog module of the ILS.

BiblioCommons involves extensive use of patron data. It does not harvest the patron records from the ILS, but as patrons register on BiblioCommons, they are validated against their patron record in the ILS. All social features of BiblioCommons can be invoked only after the patron authenticates with the library-assigned username and pin. This requirement for authentication engenders a more trusted social environment.

The interactions between the ILS and BiblioCommons take place through a Web services layer which supports synchronization of the collection data as well as real-time interaction needs for current item status and patron requests.
Like many of its competitors, BiblioCommons relies on Apache Lucene and SOLR as its core search engine technology. Other components include PostgreSQL. Most of the server internals have been implemented in Java, including the data integration and service layer. The Web application layer was developed with Ruby on Rails, a very flexible programming environment that has recently seen widespread use in the open source community. BiblioCommons implements a service-oriented architecture, with an API made available through Web services. All communication between the BiblioCommons' own Web application and the internal server applications operate through the REST (representational state transfer) API, which is also available to the library for any custom applications it may choose to implement.

Communication between the ILS and BiblioCommons takes place through a software application that resides within the library's technical infrastructure. This connector manages the transfer and synchronization of bibliographic, item-level, and authority data between the ILS and BiblioCommons. This application monitors the ILS for any changes so that data can be synchronized in as close to real time as possible. The connector also handles the interactions with the circulation module and patron data needed as users make requests through BiblioCommons that involve the local ILS. The connectors involve programming specific to each ILS product, taking advantage of any API's that might be available as well as standard library protocols. BiblioCommons has been designed to operate with any ILS, though initially the connectors have been completed for SirsiDynix Horizon and Symphony; development of connectors for Innovative's Millennium, the open source Evergreen ILS and other ILS products are underway.


In addition to the standard licensing fees associated with BiblioCommons, NYPL made a $1 million investment in the company in support of its ongoing development.

BiblioCommons offers a complete set of discovery layer services. Its implementation requires harvesting and synchronizing bibliographic and holdings information from the NYPL Millennium installation, as well as a variety of real-time functions.

The real-time interactions between BiblioCommons and Millennium take place through intercepting and parsing HTML pages as delivered through the WebPAC servers and through calls made through Patron Web Services API toolkit licensed from Innovative by NYPL.

Bibliographic data are harvested from Millennium and loaded and indexed by BiblioCommons. The search technology on BiblioCommons differs from that of Millennium, presenting differed relevancy rankings.

NYPL along with Boston Public were development partners with BiblioCommons in the creation of new functionality to fully integrate e-book discovery and lending within the discovery environment without a wholesale hand-off to Overdrive’s platform. This project was launched publicly in Jan 2012.
ReCAP requests from NYPL

Until recently patron requests for materials from the ReCAP facility have been based on handwritten paper request slips mediated by NYPL personnel.

The library has recently implemented a request form allowing patrons to make ReCAP requests online. This form has currently been implemented through the Millennium WebPAC and has not been implemented in BiblioCommons, the end-user interface used by most patrons.

The Web-based request form requires the patron to enter the barcode number from their library card and their 4-digit pin. The Web form gathers all of the data needed to submit the request to the GFA system.

The NYPL web site offers an information page for patrons explaining “Access to Offsite Collections:”

http://www.nypl.org/help/get-what-you-need/access-offsite-collections

Catalog Status: “If the item has the message ADV REQUEST, the item is in remote storage. A request for this item must be made 24-48 hours in advance. The item may not be taken from the library and it is not available for placing holds.”

A courier service makes deliveries from ReCAP to NYPL on Monday through Saturday. The Saturday service was added in recent months to improve service.

The ReCAP request form is currently available only through the WebPAC Pro interface. There are plans to eventually offer a similar request form through BiblioCommons. This task will require additional programming by BiblioCommons development personnel. BiblioCommons to date has not been willing to commit to the development of NYPL-specific customizations needed for the support of the ReCAP request form.

NYPL uses the ILLiad software from Atlas Systems to manage its interlibrary loan requests (www.atlas-systems.com/illiad). As with the other three ReCAP institutions, an ILLiad workstation is maintained at ReCAP and operated by its staff to satisfy ILL requests from its collections held in the facility. Requests for articles and book chapters are fulfilled through scanning and electronic document delivery.
The fulfillment process will need to recognize the different policies that apply to ReCAP materials. Some materials may not be available, for example, for loan through MaRLI. A broader range of materials may be requested by NYPL patrons for use on-site. A key part of the shared ReCAP collection will involve identifying what groups of materials, as identified by ReCAP Customer Codes, will be in the shared collection, and to develop a policy matrix of patron categories and customer codes that indicate authorized request scenarios.

NYPL staff indicated that the need for a more robust system for handling ReCAP requests. They report a fairly high number of requests that do not get fulfilled within expected time frames. Readers complain that they do not receive materials within 24 hours. Staff reported “ReCAP has been broken for some time now.” No specific statistics are available on the performance of the timely fulfillment of ReCAP requests, but perceptions reflect that much improvement is needed. The problems tend not to be related to the ReCAP facility and the way it process materials, but more with incomplete or inaccurate request data or in the opportunities for error in delivery and patron notification.

The problems with timely fulfillment of ReCAP requests have been attributed primarily to inaccurate information associated with the request. ReCAP can only process requests with valid data for the item, patron, and pick-up location. Requests that include invalid pick-up locations are rejected. From ReCAP’s perspective, the accuracy is 100 percent, but there are many opportunities for error outside the boundaries of ReCAP internal operations and the GFA software. From the perspective of the libraries, ReCAP requests have a lower rate of successful fulfillment than desired. The high-profile nature of the Central Library Plan increases the importance of introducing technology and processes that ensure successful fulfillment of ReCAP requests within expected timeframes and to eliminate failures.

Requests currently come in through a variety of channels, including call slips filled out by patrons in the library, e-mail requests, by telephone, and by the recently launched form integrated into WebPAC Pro. ReCAP requests are not directly managed through the holds system of Millennium. There have been ongoing problems with patrons not receiving materials within the expected time frames. Especially when the requests take place through the hand-written forms, there are many opportunities for error.

Another concern involves the expectation that high-use items will be brought back from ReCAP and returned to the on-site collection. Questions were raised about whether the concept of the shared collection supports this expectation. If an item owned by NYPL and another partner is de-duplicated within ReCAP, can NYPL still request its return?

While not directly related to ReCAP, NYPL also works with Clancy Cullen to manage the transfer of materials out of its decommissioned buildings and to stage materials out of the Stephen A. Schwarzman Building until the additional ReCAP modules are complete. A large number of materials are held in the storage facility Clancy Cullen maintains in upstate New York. Clancy Cullen also uses the GFA system to manage its inventory. When the ReCAP modules are completed, these materials will be transferred from Clancy Cullen.
Princeton University

Princeton University currently has just over 2 million items housed in ReCAP plus an additional 280,000 items from the Princeton Law Library.

With close proximity to the ReCAP facility, Princeton uses its own personnel to transport materials to and from ReCAP.

Princeton University Libraries share the vision of making ReCAP materials more easily discoverable and easily requested by the patrons associated with all the partner institutions. Princeton patrons should be able to place requests for ReCAP materials without staff mediation.

Some of the functionality that staff would expect would include tools that help with the process of de-duplication of its materials relative to the ReCAP collection. This would apply retrospectively for materials already owned that might be duplicated in ReCAP. More importantly, prospectively there is an interest in developing collections in consideration of the materials in ReCAP to avoid unintentional duplication. In addition to any patron discovery tools, there should also be a staff interface developed to support selectors in making decisions about new acquisitions, transfers to ReCAP, or withdrawals.

Other information mentioned that would be helpful to selectors would include availability of materials in HathiTrust, through the Center for Research Libraries.

Both for library patrons and for staff use, there is interest in a federated index of the bibliographic information of the three institution’s ReCAP holdings. A federated index of the three institution’s entire collections would be useful for collection development, but may not be within the scope of the project.

Automation strategy

Princeton University currently uses the Voyager integrated library system from Ex Libris. This system has been in place since about 2000 when it converted from NOTIS.

The Voyager ILS includes a variety of APIs that enable its integration with external systems.

According to Duong the WebVoyage Web Services design is characterized by stateless, atomic transactions and consumer/provider (client/server) architecture [1]. Each web service requires an OPAC instance to access the Oracle database. Ex Libris designed these Web Services for the following rationales: the interoperability between different systems, the integration between different applications, to enrich customer functionality, and to create a federation of resources and to facilitate faster time for software development cycle. Detailed Information regarding these services is outlined in the collaborative Web site: http://www.exlibrisgroup.org/display/VoyagerOI/XML+Over+HTTP+Web+Services. Web Services include search services, media services, patron requests services, etc. Duong (2009) further identified the process of setting up the web services. The configuration file on Voyager server is vxws.ini. OPAC/CIRC server configuration in the server.xml was also verified. And further Apache Web Server, Tomcat, and Oracle database connection were validated, and that OPACSVR/CIRCSVR is operational.

from this meeting are available at:
http://www.exlibrisgroup.org/display/presentations/Developer+Meets+Developer+March+2009

[Ho, Birong. “The Integrated Library System’s APIs, an Open-source Web 2.0 Catalog, and University Computing Live Happily Ever After.” Code4lib Journal 12, 2010-12-21
http://journal.code4lib.org/articles/4165]

Alma and implementation status at Princeton
Princeton University Libraries has been working with Ex Libris as a development partner for Alma, the company’s new generation library management platform. As a development partner, Princeton has been able to work with each of the incremental releases of the software, testing functionality, with opportunities for input on the development of the product. Other development partners include Purdue University, Boston College, and KU Leuven. Boston College became the first library to place Alma into production use on July 2, 2012.

Princeton was represented on the Expert Advisory Group to advise Ex Libris on the Community Catalog of Alma by Katharine Treptow Farrell, Head, Order Division, and Assistant University Librarian for Technical Services.

The following excerpt provides an overview of Alma:

Alma aims to strike a balance between providing automation services through shared resources and those that are managed locally. Its conceptual approach includes multiple layers, or zones:

- The Community Zone provides a layer of broadly shared services. The community zone will include a large metadata store of bibliographic that Ex Libris will initially provide accessible by all Alma libraries. This store will also grow through contributions made by libraries as they create new records. Records in the community zone can be enhanced as needed by libraries using Alma.
- The Library Zone is a metadata store specific to a single organization implementing Alma. Libraries may have various circumstances which might lead them to maintain some of their metadata in a private way.

A library’s holdings are managed through another data store called the Inventory, which can associate with descriptive data from either the Community or Library zones. As libraries add new items to their collections, they can tag onto existing records in the Community Zone, create new records in either zone as needed, and can retrieve records from external bibliographic services as needed, which can then be contributed to either zone, depending on concerns like record use restrictions. Ex Libris intends the metadata in the Community Zone to be available openly.

The global services included with Alma will allow libraries to leverage widely shared resources. The bibliographic records available in the Community zone will facilitate efficient workflows when adding new materials to their collections. One of the broad goals of Alma is to allow libraries to execute common tasks in the most efficient ways so that they can concentrate more on the unique activities of more interest and strategic value to their organizations.

The Metadata Management System provides tools and services related to metadata. This component will include editors and other tools that are involved in the creation and maintenance of
metadata. ALMA will take a format agnostic approach to metadata. Since it provides services for all types of materials, it will not be tied to MARC, but will be able to provide editing and validation tools for all applicable formats. The Metadata Management System will also support management of licensed electronic resources, relying on the knowledgebase currently associated with its SFX link server.

Ex Libris will offer Alma through software-as-a-service, hosted in a cloud infrastructure. During the development phase the system is hosted in Amazon's Elastic Compute Cloud. The company has not yet determined whether the production product will be delivered through the Amazon EC2 platform or some other competing infrastructure. The cloud infrastructure that Ex Libris ultimately selects will be based on cost, performance, and service-level agreements. While it's important to understand that Alma has been designed for deployment through the cloud, the specific provider ultimately selected will not necessarily be apparent to the libraries that use the product. [Ex Libris has since announced they will be using Equinix as its cloud infrastructure provider.]

Ex Libris has indicated that it intends to support local installations of Alma. Large consortia, for example, may prefer to implement and manage their own instance of URM. While the product has been designed for delivery through SaaS, some libraries may have specific needs that require a local installation.

The staff interfaces for Alma will be entirely Web-based. No local software will be required for library personnel to operate Alma and it is expected to function will all major Web browsers.


Although Princeton University has been working with Ex Libris for Alma, it has not set a specific date by which it plans to implement Alma as its production environment. The library plans to wait until it considers the software finished. It has developed a list of functionality that must be in place before going live with the software.

Informally, a transition to Alma might take place in December of 2012 or in July 2013 depending on progress made by those dates.

Depending on the dates of implementation of the technology for the next phase ReCAP, there should be a short-term plan to build interfaces into the current Voyager implementation as well as longer-term integration with Alma.

Ex Libris emphasizes that Alma will expose a robust set of APIs in addition to the functionality delivered through its staff and public interfaces. These APIs seem very well suited to the kinds of interoperability that will be needed to connect systems together more seamlessly in the technology to support the next phase of ReCAP.

Princeton University staff report, however, that so far in their experience with Alma as development partners that they have not had the opportunity to work with its APIs. The focus of effort has been on the delivered functionality. Any work with the APIs would have helped to reinforce their use by the
larger ReCAP project. The availability of the APIs will need to be further verified, including which APIs will be exposed in what releases of Alma.

**Discovery Strategy**
Princeton University currently has a variety of end-user interface and discovery tools, including the native WebVoyage online catalog delivered with Voyager, Primo, and Summon.

WebVoyage continues to be offered and is used by those with advanced research needs. It offers additional search options, including the ability to browse the collection based on authority headings or call numbers. Princeton does not plan to withdraw this interface until similar capabilities are available in its other discovery services.

Princeton has also implemented the Primo discovery product from Ex Libris. The local Primo index is currently populated primary with the records from Voyager. Records exported from Voyager are processed through a process, called a “pipe” in Ex Libris terminology, that normalizes the data as it is indexed. The Primo index is updated three times a day with added, modified, and deleted items.

Primo displays the current status of materials through a real-time query of the underlying management system, which is Voyager in the case of Princeton. This real-time availability is accomplished in one of two ways. One alternative uses a Web service built into Voyager. This service does not work well for some items, such as those that have hundreds of associated items. Princeton has also developed a custom program that retrieves the MARC, holdings, and item records associated with a title which works reliably even for those with extensive holdings data.

Patrons can choose to use Voyager catalog or Primo from the library’s main Web page. No specific statistics have been provided, but the perceptions are that use of Primo is increasing relative to WebVoyage.

Princeton has not implemented Primo Central from Ex Libris which provides a deep index of articles, book chapters, and other resources, profiled according to the library’s holdings. At the time that the library was ready to provide this kind of service, it determined that Summon provided more comprehensive coverage of its subscriptions to electronic resources.

Summon from Serials Solutions is offered as a separate discovery tool for articles. Materials from Princeton’s Voyager ILS are not loaded into Summon, as do libraries that intend to use it as their comprehensive discovery service. Princeton currently positions Summon primarily as discovery tool specialized for e-journal articles and other electronic materials. Since it does not include books and other local materials, it does not need to interface directly with Voyager.

Princeton has developed an online form which has been integrated into both WebVoyage and Primo for ReCAP requests. The form is offered for materials coded with ReCAP locations. The form collects patron and item information, and lists the valid pick-up locations available. The form does not make direct updates within Voyager to the item status or patron record.
Data created the ReCAP request form is formatted according to the specifications developed by partner institutions and transferred by FTP to the GFA system three times per day. In the past each request was transmitted instantly, but was changed to scheduled batch transfers since some requests were lost. The scheduling of the batch transfer is scheduled to be available to ReCAP staff as they generate the pick lists each day.

Princeton does not have workflow management tools in place to manage the handling of ReCAP requests and returns. When a request is made that cannot be fulfilled by ReCAP, an error report is generated and transmitted to the library. These error reports are reviewed by library staff. In some cases the problem relates to requesting items that are not actually available to that patron or for the requested pick-up location, or that the item number does not exist in ReCAP, or that it has since been requested by another patron. The batch-oriented nature of the current processes lead to many opportunities for items to appear as available when they are not. Having more dynamic status information as materials flow in and out of ReCAP would address this issue.

One of the most common errors that Princeton reports in ReCAP requests involves invalid pick-up location. Any new technology tools created should fully validate all parameters of the request.

The request process does not automatically inject the request into Voyager. The request does not appear in the patron’s account and the item status remains as available. As items are received from ReCAP, a staff member will manually place the item on hold for the patron account based on information on the printed ReCAP slip. Once placed on hold, a notice is sent to the patron. Items are then charged through normal circulation procedures.

The patron authentication environment at Princeton University primarily relies on barcode number and pin (last name) login credentials held in Voyager. This sign-in process is used by patrons as they access their account either through WebVoyage or Primo. The campus maintains a LDAP-based authentication service which is used by the library for Borrow Direct requests. Princeton has not yet implemented a single sign-on environment that spans all the services available to students, faculty, or staff. A Shibboleth service is available but has not been implemented extensively. The HathiTrust is one example of a resource accessed via Shibboleth.

Final decisions have not been made regarding how Alma might fit into the campus-wide authentication environment. Initially the implementation would be based on similar barcode and pin authorization as is now done in Voyager.

Summon does not require authentication for search. Access to items linked to in Summon depends on IP authentication, based on physical presence on campus, remote VPN access, or through a proxy service. Display of full-text subscribed resources requires authentication.

Princeton uses SFX and its associated knowledge base as its primary link resolver. The operation of Summon requires using the 360 Core from Serials Solutions to select the resources to be activated in the index. This arrangement requires keeping the 360 Core knowledge base (KnowledgeWorks) in synch with the SFX Global KnowledgeBase to the fullest extent possible. Princeton staff have learned that
packages and portfolios may often be organized differently in these two products, making it difficult exactly match all resources.

Princeton selected Summon because it determined that it more fully represented its electronic holdings. Primo Central has expanded its coverage since that decision was made. At some point Princeton may revisit its discovery options, but for the time being they plan to maintain this hybrid system.

New capabilities for the ReCAP will need to be created primarily with the Primo interface in mind, but should also take into consideration the expected ongoing use of WebVoyage and Summon.

Primo includes a capability Ex Libris calls “Deep Search” formerly referred to as “Third Node” that integrates access to resources that have compatible index structures to achieve immediate search capabilities without having to ingest the individual items in to the local instance of Primo. This capability works only for resources. Additional investigation will need to be performed to determine whether this capability can be anticipated to work well for blending an index of the shared ReCAP into Princeton’s Primo instance.

One alternative to integrating the ReCAP Shared Collection into discovery layers such as Primo would be to export records form the ILS of the partner institutions.

The ETH-Bibliothek in Zurich has created a “Primo CMS Colr Websearch Adapter.” That enables a Primo installation to connect directly to any SOLR server and directly integrate content into Primo.

Columbia University Libraries
Columbia has approximately 3.7 million items currently stored in ReCAP.

Automation Strategy
Columbia University, like Princeton, uses the Voyager ILS as its primary automation environment. These two university library systems have entirely separate instances of Voyager. Columbia has been using Voyager since 2002 when it migrated from NOTIS.

Columbia University was a participant in the Kuali OLE project (kuali.org/ole) to create on open source enterprise-ready library management environment for research libraries during its initial 1-year planning phase (2008-2009). It did not join the subsequent 2-year project to build and implement the software.

2CUL
Columbia University and Cornell University have engaged in a partnership called 2CUL to collaborate which will include a “broad integration of resources collections, services, and expertise” (2cul.org). One aspect of this collaboration will include the implementation of a shared library management environment, with Alma as the primary system under consideration. 2CUL brings Columbia and Cornell into a close partnership while still maintaining their separate organizational structure and identities. The partnership will result in collaborative collection development, shared technical processing, and reciprocal access to collections by their respective patrons. 2CUL has received support from the Andrew W. Mellon Foundation.
Columbia’s involvement with 2CUL has implications for the ReCAP project. Should 2CUL result in an automation environment shared between Columbia and Cornell, details will need to be worked out relative to Columbia’s relationships to ReCAP that may not apply to Cornell.

Given that the commitments for Columbia to migrate from Voyager to Alma are not firm, it should be assumed that short term development should target the existing Voyager and Blacklight environment.

**Authentication Environment**
Columbia University maintains a central authentication service. The University has created an Apache module called mod_auth_pamacea to implement PAM (pluggable authentication modules) authentication. The Columbia University Libraries use PAM in conjunction with the EZ Proxy from OCLC to perform authentication for restricted resources and services.


**Discovery Environment**
Columbia University has developed a discovery interface based on Blacklight. Blacklight is an open source discovery interface based on the Ruby on Rails development environment and the Apache SOLR search technology (projectblacklight.org). Original development work for Blacklight was initiated at the University of Virginia and it has been adopted by other major institutions such as Stanford University, the University of Wisconsin, Johns Hopkins University, North Carolina State University, Indiana University, the University of Hull in the UK, and by WGBH in Boston.

The beta version of Columbia’s CLIO Blacklight catalog has been made public since March 2012: [http://cliobeta.columbia.edu/](http://cliobeta.columbia.edu/)

CLIO Beta follows a segmented approach to delivering search results. Parallel listings are provided for the 7.7 million volumes managed in Voyager and for those that come from Summon, the Academic Commons and other sources.

Columbia uses the 360 Suite of products from Serials Solutions for management and access to electronic resources.

Columbia University began subscribing to the Summon discovery service from Serials Solutions in 2010. It uses the company’s 360 Core to profile its e-resource holdings and 360 Link as its OpenURL link resolver.

**Resource Sharing Programs and Services**
Columbia participates in Borrow Direct, which uses interlibrary loan management software from Relais International and the MasterKey federated search technology Index Data as the basis of an expedited book request and delivery service. Other participants in Borrow Direct include: Brown, Columbia, Cornell, Dartmouth, University of Pennsylvania, Princeton and Yale. Borrow Direct has been in operation since 1999. It was originally implemented using the Ursa system from SirsiDynix, which has since been discontinued. The Borrow Direct system communicates with the ILS of each participating

Columbia University offers a **Faculty Document Delivery Service** that allows faculty affiliated with the Morningside Heights Campus, Barnard College, and Union Theological Seminar to request articles held in the Butler Library to be scanned and delivered electronically.

Columbia participates in **MaRLI**, which gives selected faculty and graduate students access to collections of NYPL. This arrangement is done within the existing borrowing environments of the three institutions and does not involve any additional supporting technology.

Columbia also participates in the **RapidILL** service (rapidill.org) for resource sharing of articles. RapidILL was developed and is managed by Colorado State University. RapidILL was initially implemented at Columbia University as a centralized service, but it is being phased onto a more decentralized operation.

The [Avery Architectural and Fine Arts Library](http://www.library.columbia.edu/av-ill/ill/illiad) has an electronic document delivery service that uses the ILLiad interlibrary loan management software from Atlas Systems.

Columbia University became a partner in the Google Books mass digitization project (announced December 2007) and joined the HathiTrust shared digital repository in December 2009.

**ReCAP requests**

Request from Columbia Patrons for recap materials take place primarily through an unmediated process using forms presented through the online catalog and discovery interfaces. Consistent with the processes implemented by the other ReCAP institutions, a Web form in the online catalog or discovery interface produces text file with the data representing the item and patron. Request data are aggregated into a text file that is then transmitted at specified times to the ReCAP GFA system.

A form for patrons to request off-site materials at ReCAP has been implemented in both the native Voyager catalog and in Blacklight CLIO. Request services have also been implemented for interlibrary loan and Borrow Direct.
The processing of ReCAP materials to fulfill patron requests for materials in ReCAP is handled by the various circulation service points among the Columbia University Libraries. These service points serve as pick-up locations for ReCAP materials. Interlibrary loan, in contrast, is operated more as a centralized service.

**ReCAP Issues identified by Columbia**

Columbia University is interested in a more seamless approach for borrowing ReCAP materials from the other partner institutions. The current approach works more like ILL. Current methods and procedures are focused on retrieving their own material from the facility, with no easy way for patrons to discover and request materials from the other two institutions.

Columbia personnel characterized the current ReCAP procedures as delicate and that they don’t always work well. The batch processes involved for transmitting ReCAP requests have problems. The current processes do not accurately reflect what items are actually available. It is important that the new system be able to handle multiple simultaneous requests, which is not possible in the current batch process.

As general background, Princeton personnel report that their faculty members continue to be uneasy about ReCAP. Faculty lament that they cannot browse the stacks for all library materials. ReCAP must have excellent performance in its service to help library customers accept this arrangement. Service expectations will only increase over time. Princeton wants to promote ReCAP not only as increasing the amount of collection materials available to their patrons, but also as an improvement in service.

Columbia expects the new system to be very intelligent about the way that it exposes the shared ReCAP materials through its discovery interfaces.
Part II. Technology Implementation Options

This section provides an overview of concepts and issues relevant to the next phase of the technology support for ReCAP. It outlines some of the general requirements needed to fulfill the project goals related to management and access of a shared ReCAP collection.

Discovery of ReCAP Materials

In order to support the concept of a shared ReCAP collection with the ability for patrons to directly request materials, enhancements will need to be made to the discovery interfaces and online catalogs maintained by the three consortium members.

Key facts and assumptions:

- NYPL, Columbia, and Princeton University each use different automation systems and discovery services.
- There is no expectation that the three institutions would converge on a single discovery environment.
- End-user access to the ReCAP materials should not be offered through a separate interface.
- Materials available to their patrons through the shared ReCAP collection should be available through each institutions existing discovery interface, or through new interfaces that may be implemented in the future.

The Library Archival System developed by Generation Fifth Applications operates as an efficient inventory management system and does not include descriptive bibliographic data. One of the key considerations for the new ReCAP technology environment will be to not disrupt the existing operational efficiencies of LAS as it is used to manage the ReCAP facility.

Bibliographic data related to ReCAP materials resides only in the automation systems of the three partner institutions. The barcode number on an item provides the key linkage between bibliographic data and the ReCAP inventory.

Status and availability of materials in ReCAP is maintained in each ILS. Holding codes within each partner ILS are changed to ReCAP locations. ReCAP Customer Codes may or may not be stored in the ILS.

Status and availability of materials is maintained in the ReCAP GFA System. No systematic synchronization is performed between item status in GFA and the ILS of the three participants. In the current environment, the status of an item remains as available between the time it was requested and when it is pulled and shipped to the library.

Participants have indicated that they would prefer not to have to load the MARC records representing the holdings of the other ReCAP institutions into their local automation systems. Cross loading MARC records across partner ILS implementations would introduce significant costs and support burdens.
some cases current system capacity or license thresholds would be exceeded. While cross loading MARC records is presented as a theoretical option, it is not a desirable alternative.

The ILS of the owning institution does not currently maintain real-time availability of an item throughout the lifecycle of the request, fulfillment and return process. Items continue to show as available between the time a patron request is made and when it is processed in the library for patron pick-up. There are significant intervals when even the owning library’s ILS shows items as available when they are not.

A discovery infrastructure will be created to provide improved patron access to the shared ReCAP materials. At a minimum, this discovery component will be populated with bibliographic and administrative metadata for the items identified as belonging to the shared ReCAP collection.

Option: Should the discovery environment also be populated with a broader selection of ReCAP records? Index views could be defined to scope search to the Shared ReCAP collection for patron search and other views defined for library personnel to support collection management activities. The design of the new infrastructure should accommodate multiple scenarios for what records are harvested and indexed.

Is there a need to filter inclusion of ReCAP materials by institution? Are there examples where materials should display in the ReCAP index of the home institution and not the partner institutions?

Should there be a status flag in the patron records for MaRLI which expands a patron’s eligibility for ReCAP materials? What other kinds of patron scenarios need to be accommodated in the ReCAP discovery and delivery environment?

**Expected functionality for ReCAP Discovery**

Discovery of ReCAP materials should not require users to go to a separate service. New functionality will be created that allows patrons to see the shared ReCAP materials among search results and to place requests for them. Authentication of patrons will rely on mechanisms currently in place.

Discovery of ReCAP materials should be integrated into each the discovery services or online catalogs in general use for the participating institutions.

- **Issue:** The creation of an index for ReCAP materials could likely be integrated into discovery layer products such as Blacklight, Primo, or BiblioCommons designed for segmented index architectures. Integrating this capability into online catalogs will be much more of a challenge without having to resort to loading the MARC records for the shared ReCAP collection into the respective ILS instances.

ReCAP materials for each library’s home institution will already be present in their ILS which directly drives its online catalog and is used to populate its discovery layer.

The key extension for the next-phase ReCAP involves the inclusion of materials from the other two institutions.
Search results should appropriately label materials as “off-site” and maintain internally the GFA Customer Codes or other factors needed to calculate eligibility for request for each request context and valid pick-up locations.

New system should provide a service that provides accurate data in real time regarding whether an item is available for request. This real-time availability status service would be consumed by the discovery or online catalogs of the partner institutions and may be useful to other business processes.

ReCAP materials subject for discovery will include all types of materials: monographs, serials, media, etc.

The infrastructure should be flexible regarding how items are defined as belonging to the ReCAP Shared Collection. Possible factors for inclusion criteria might include GFA Customer Codes, ILS holding locations, item types, date ranges, or designated individual items.

At this phase recap discovery will include journals at the title and issue level. It will not directly include article-level data.

Any article-level discovery products used by the partner institutions should be able to appropriate link to items in the Shared ReCAP collection should be handled through OpenURL link resolvers or other mechanisms. Resolver menus would offer electronic document delivery options for these materials.

Within each discovery environment, ReCAP materials should be capable of being interfiled among search result listings or offered as a separate result set, depending on library preference or by the user through facet selection.

Through what interfaces is discovery expected for the Shared ReCAP collection:

- NYPL WebPAC Pro online catalog
- NYPL BiblioCommons discovery service
- Princeton University Main Catalog (WebVoyage)
- Princeton University SearchIt@PUL (Primo)
- Columbia University Classic Clio Catalog (WebVoyage)
- Columbia University CLIObeta (Blacklight)

ReCAP results will not be included in:

- Princeton University Articles + (Summon core index only)

*Technical implementation options*

**Cross-loading Records in each ILS**
One approach that would be the most technically feasible but include the most overhead would involve loading the records from the partner institutions into each other’s ILS implementations:

- Load Princeton and Columbia sharable ReCAP records into the Millennium ILS of NYPL
- Load NYPL and Columbia sharable ReCAP records into the Voyager ILS of Princeton
• Load NYPL and Princeton sharable ReCAP records into the Voyager ILS of Columbia

This approach has disadvantages:

• It would introduce millions of records into each automation environment which would need to be maintained.
• It could greatly increase ILS licensing costs since fees include as a factor the number of bibliographic records managed.
• It would greatly increase technical services processing needed to maintain authority control and other database quality assurance issues.
• It would introduce significant work to reconcile serials holdings represented in the ILS based on materials in the other ReCAP institutions.

Support of this model would include several procedural steps and technology components:

• Selective extraction of records from each of the three ILS implementations
  o Extraction would include full MARC records with holding and status information encoded on 9xx fields.
  o Current status data will be included on the extraction and data loading routines, but will be definitively displayed based on the new real-time availability service.
  o Selection of records to be extracted would be based on criteria defined by the ReCAP participants. Factors for the criteria might include: ReCAP Customer Code, ILS holding location, Item Type, Date Range, or individually designated items.
    ▪ Options for selection would include whether to include all ReCAP customer codes or just those designated as sharable by other partners.
  o Extraction would be performed through standard procedures as described in the DLF Integrated Library System – Discovery Protocol (ILS-DI), which generally corresponded to the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).
  o Extracted records would be held in a common file system for later processing.
  o Incremental extracts would be performed at least daily for new items transferred to ReCAP, for withdrawals, or of other changes.
  o Supporting technology components might include:
    ▪ OAI Toolkit from the eXtensible Catalog project
      “The Metadata Services Toolkit enables the XC user interface to present FRBRized, faceted navigation across a range of library resources. The toolkit aggregates metadata from various silos, normalizes (cleans-up) metadata of varying levels of quality, and transform MARC and DC metadata into a consistent format for use in the discovery layer.”
      http://www.extensiblecatalog.org/
  o Extracted records would also need to be extended with date derived from the GFA Library Archival System.
    ▪ Based on the barcode number as the unique match point corresponding records would be retrieved from GFA.
- Customer Code (designates owing library and collection and used to determine policies on eligibility for request, loan periods, etc.)
- Status information – availability
- These data points would be encoded on the 9xx fields since they would be needed to support a preliminary status display in discovery systems and request eligibility. [Are ReCAP customer codes already present in ILS bibliographic records? Can they be definitively calculated? It would simplify the process if retrieval of GFA records could be eliminated.]
  - Record Extraction routines will also be needed for other implementation scenarios
  - Loading of MARC records into ILS systems
    - Records already exist for ReCAP holdings of each institution
    - Records would be loaded into each of the three ILS implementations of the MARC record sets from the other two institutions
    - The process would use the standard record loaders that parse the 9xx fields to create holdings and item records
    - Depending on the cataloging standards of each of the institutions, records may need to be processed to validate headings or other tasks needed to preserve the integrity of authority control
    - Possible processing required to deal with serial records and holdings
    - Processing should detect duplicates and carry out selected processing options:
      - Suppress duplicate records and add additional copy statements to existing bibliographic records
      - Reject and issue error report
      - Load duplicate record and generate informational message
    - Incremental loads would be performed at least daily to synchronize ILS holdings with ReCAP inventory across participating institutions
  - Enable Real-time availability
    - The creation of a real-time availability process will be needed by any of the implementation scenarios and is described below.
Creation of a modular ReCAP index or plug-in
This approach would involve the creation of an index that could be integrated into the relevant discovery services.

A discovery environment would be based on an index created from bibliographic records representing the Shared ReCAP collection. It would rely on a bibliographic database as a component of the ReCAP infrastructure created according basically the same requirements as those used for the bibliographic cross-loading option. The key difference is that the bibliographic database remains in the central ReCAP application and no additional records are loaded in the partner institution ILS implementations.

- Selective extraction of records from each of the three ILS implementations
  - Extraction would include full MARC records with holding and status information encoded on 9xx fields
Current status data will be included on the extraction and data loading routines, but will be definitively displayed based on the new real-time availability service.

Selection of records to be extracted would be based on criteria defined by the ReCAP participants. Factors for the criteria might include: ReCAP Customer Code, ILS holding location, Item Type, Date Range, or individually designated items.

- Options for selection would include whether to include all ReCAP customer codes or just those designated as sharable by other partners.

Extraction would be performed though standard procedures as described in the DLF Integrated Library System – Discovery Protocol (ILS-DI), which generally corresponded to the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).

Extracted records would be held in a common file system for later processing.

Incremental extracts would be performed at least daily for new items transferred to ReCAP, for withdrawals, or of other changes.

Supporting technology components might include:

- OAI Toolkit from the eXtensible Catalog project
  
  “The Metadata Services Toolkit enables the XC user interface to present FRBRized, faceted navigation across a range of library resources. The toolkit aggregates metadata from various silos, normalizes (cleans-up) metadata of varying levels of quality, and transform MARC and DC metadata into a consistent format for use in the discovery layer.”
  

Extracted records would also need to be extended with date derived from the GFA Library Archival System.

- Based on the barcode number as the unique match point corresponding records would be retrieved from GFA
- Customer Code (designates owing library and collection and used to determine policies on eligibility for request, loan periods, etc.)
- Status information – availability
- These data points would be encoded on the 9xx fields since they would be needed to support a preliminary status display in discovery systems and request eligibility. [Are ReCAP customer codes already present in ILS bibliographic records? Can they be definitively calculated? It would simplify the process if retrieval of GFA records could be eliminated.]

Record Extraction routines will also be needed for other implementation scenarios.
Apache SOLR / SOLRMARC

The obvious candidate technology to support this approach would be the SOLR and Lucene search and retrieval applications available as open source under the governance of the Apache Foundation.

One implementation option would involve using the general infrastructure of one of the SOLR-based discovery environments to provide a platform for creating indexes and for metadata management tasks. Options include:

- **Blacklight**: open source discovery services based on Apache SOLR Lucene and built using the Ruby on Rails development environment
- **VuFind**: open source discovery service based on Apache SOLR Lucene and built using the PHP programming language
One instance of this platform would be maintained on behalf of all the ReCAP institutions. The platform would need to be sized to accommodate indexing processes and other support tasks at the scale of the combined ReCAP inventory.

The SOLR index would be created by indexing the bibliographic records harvested and synchronized from the partner ILS implementation. Data would staged in the MARC-based bibliographic database and SOLRMARC used to build and maintain the index accessed by discovery systems.

It is also possible to associate the bibliographic database with a Z39.50 or SRU service to allow integration of the Shared ReCAP Collection by systems not compatible with Apache SOLR or Lucene.

Figure 10 Figure 10 Creation of SOLR index derived from central MARC database
Additional metadata support would be provided through the tools developed through the eXtensible Catalog project:

- **XC NCIP Toolkit**
- **XC Metadata Management Toolkit:**

  The Metadata Services Toolkit enables the XC user interface to present FRBRized, faceted navigation across a range of library resources. The toolkit aggregates metadata from various silos, normalizes (cleans-up) metadata of varying levels of quality, and transform MARC and DC metadata into a consistent format for use in the discovery layer.

  This toolkit presents an opportunity for libraries to apply their expertise with creating and managing metadata in a variety of web applications.


Data exchange and communications capabilities for the environment might include components such as:
http://code.google.com/p/xncip2toolkit/wiki/VoyagerInstallation
- Patron Account Information API
http://gbv.github.com/paia/paia-868166f.html
- DAIA - Document Availability Information API
http://www.gbv.de/wikis/cls/DAIA_-Document_Availability_Information_API
“The Document Availability Information API (DAIA) defines a data model with serializations in JSON and XML to encode information about the current availability of documents. This document defines the serialization formats DAIA/JSON and DAIA/XML and a HTTP query API to query DAIA information.”

This approach will be technically challenging. It assumes an environment of discovery services capable of integrating with third-party indexes in addition to the one built-in to the service. Some of the considerations of this approach:

- Bibliographic records corresponding to each institution’s sharable ReCAP records will be extracted from the respective ILS implementations as described above
- Records will be indexed in SOLR
- Facets or other mechanisms will be implemented to allow segmentation of the index according to the three owning institutions
- The environment will need to be capable of presenting indexing views based on selected institutional segments. The ReCAP SOLR index to be integrated should activate only the ReCAP items from the other two institutions and not duplicate indexing already present in their discovery environment for their own ReCAP items
- This approach will be challenging to include in the native online catalogs for the three institutions since they do not make use of discovery indexes but rather are generated directly out of the respective ILS implementations.
- All three discovery environments are based on SOLR:
  - NYPL BiblioCommons
  - Columbia’s Blacklight-based CLIO beta
  - Princeton’s Primo: based on Lucene. The Ex Libris “Deep Search” technology provides support for access to third party indexes in addition to records loaded into Primo through pipes. Primo Central, for example, integrates article-level materials with the local index using this Deep Search
- Additional investigation will need to be performed to ensure the feasibility of this index integration approach.
- If successful, multiple indexes within a discovery environment will work together to produce blended results ordered according to relevancy algorithms applied consistently among the sources. Index segments would include:
• Primary discovery index derived from materials managed through the ILS implementation
• Additional index representing ReCAP materials
• Additional indexes representing article-level materials, such as Summon or Primo Central.

• Item display routines would need to interact with different business systems for real-time availability and status:
  o For items managed within the local ILS, the standard availability routines would apply. Each of the current discovery implementations includes dynamic interrogation of the ILS for circulation status and for patron My Account functionality
  o For items representing ReCAP items from partner institutions, item display (brief or full) would trigger a call to the Web service on the ReCAP GFA system to retrieve current status
  o The status of an item in GFA LAS would be definitive only if the status is updated at the initiation of a patron request. A two-factor availability status optionally could be implemented based on both LAS status and that from the relevant ILS.

• Request forms, described in more detail below, would be presented for ReCAP items. Signed-in users might trigger additional processing that validates eligibility for request.

Creation of ReCAP Requests
A standard request form should be designed that collects consistent information and that validates all data collected. Examples:

• Authenticate the patron through available services:
  o Through the proprietary sign-on mechanism provided through the online catalog or discovery interface
  o Through NCIP call into the ILS
  o Through a campus authentication service

• Reliable performance of the fulfillment requires that patrons be authenticated and that any required details are present in their patron record.
• Ensure that patron is eligible to make request
• Patron data either derived from or validated against patron record in home institution ILS
• Validate that the pick-up location is valid for that type of material and that type of user
• Validate that they item requested is actually available through dynamic status information in ILS and GFA
• The form should accommodate requests made directly by patrons through library discovery interfaces or online catalogs
• The form should also be designed to accommodate requests mediated through library staff members on behalf of patrons. This form would be used to enter requests made over the phone, send in by e-mail, etc.
Staff-mediated requests would perform the same patron validation as when unmediated. The staff member might, for example, request the patron’s patron ID number and the system would dynamically interrogate the ILS using NCIP or other applicable protocol, to retrieve the patron record and validate request conditions.

- ReCAP requests once validated by the patron would then be submitted to a workflow management application as described below.
- Automatically transmit the request to the GFA request queue
- Update GFA system dynamically upon successful processing of transaction to flag item as in use.
- Upon successful submission of the request, the confirmation message will be displayed to the patron and optionally transmitted by e-mail
- The expected time of fulfillment should be displayed, based on when the request was submitted relative to the current schedule of ReCAP for pulling requests, courier schedules, library opening hours, and other factors.
- A transaction identifier assigned so that patrons and library or ReCAP personnel can check the status of the request throughout the fulfillment process.

**Generalized Request Process**

![Diagram of Proposed model for generalized ReCAP request processing]

*Figure 12 Proposed model for generalized ReCAP request processing*
Real-time Availability Environment for ReCAP

One of the key requirements for the next generation of ReCAP technology support will involve the ability to display the status and availability of ReCAP item in real time. Some of the problems with the existing processes have been identified as related to the batch processes used for item requests and status updates. Also in order to extend the operational scenario to include requests across all participating institutions, support for real-time status information and the ability to dynamically update statuses based on current transactions will be essential.

Some of the options to implement a dynamic status environment might include:

Implement a layer of services that dynamically interacts with all four business applications:

- The GFA Library Archival System that manages the ReCAP inventory
- The Millennium ILS of NYPL or future automation platform
- The Voyager ILS of Princeton University or future automation platform
- The Voyager ILS of Columbia University or future automation platform

Add a services layer to the GFA Library Archival System that exposes a set of Application Programming Interfaces (APIs) preferably in the form of secure REST services or enveloped in SOAP.

- This service layer would need to be created by developers with proprietary access to the GFA ToolPro environment.
  - Such work would likely be accomplished as a sponsored development project with Generation Fifth Associates, but there might be other possible scenarios
  - Access to services would require authentication to ensure system integrity
- Examples of services needed to support ReCAP might include
  - Request item status: presentation of an item identifier (barcode number) would return the items current status and availability
  - Update status: presentation of an item identifier and a desired status (available, in process, in transit, charged to user, etc.) would return an error or success flag
- This service layer could also take the form of a NCIP responder. Selected NCIP operations could be implemented with the specific functions needed to maintain a dynamic status and availability environment. This approach could take advantage of the multiple open source implementations of NCIP available, including those offered through the eXtensible Catalog project.

Create functionality that can be embedded in each of the online catalogs or discovery layers that dynamically calculates the status and availability of ReCAP items that appear in search results

- The items will include location codes that cause display of something like “off-site”
- Items will also include access to the ReCAP Customer Codes to support calculation of eligibility for request and to validate pick-up locations
- A preliminary availability status might be presented based on statuses harvested from ReCAP records and updated through the current batch system
- Appearance of a ReCAP item during the presentation of a brief results list or on a full record display would trigger a call to the ReCAP server’s availability service. This request might be implemented through AJAX (Asynchronous JavaScript and XML).
- This service could also be implemented.

**Real-time Status Polling**

**Figure 13** Single factor Model of Possible real-time status display
Management of ReCAP Requests

The improvement of fulfillment services related to ReCAP requests could be accomplished through an workflow management system that automates, validates, and verifies all aspects of the transactions involving ReCAP materials that transpire in the gaps between the ReCAP GFA system and the respective ILS implementations. The workflow management system would also provide supporting infrastructure for GFA and the ILS implementations that might not be present natively.

The GFA system excels at ensuring that no items are lost within the ReCAP storage facility. It does not perform any tracking of items as they proceed through external business processes: transit to library, processing in library, check-out to patron, routing within library branches, and transit back to ReCAP. In order to gain better control of ReCAP materials, a complete end-to-end management system is required.

The Workflow Management Utility would be a central point for working with policies that apply to ReCAP materials. It would maintain a table of policies that includes a matrix of GFA Customer Codes, ILS patron types and eligibility, valid pickup locations, as well as parameters needed in support of electronic document delivery.
A fulfillment module could optionally be extended to other types of patron services beyond ReCAP requests. Such an generalized utility would bring consistency and control to a wide range of patron requests, including holds and recalls, interlibrary loan, document delivery, in addition to ReCAP requests. This system would provide a consistent authentication and present a consistent user interface and take advantage of business logic that would present the most efficient options available to the patron for receiving a desired item, and would pre-validate all data to ensure successful fulfillment. Such an extension of capability would be outside the scope of the initial project, but it should be designed in ways to support workflows other than ReCAP requests.

Many features would be implemented as a Web service that could be addressed by any authorized discovery service of the participating institutions.

The Workflow Management Utility would maintain a transaction record for each request. Transactions would be opened when submitted through the ReCAP request form and would be closed when items are returned to the ReCAP storage modules.

Transactions would be retained in ways that preserve patron privacy but that support full reporting capabilities, including queries of pending transactions and retrospective performance statistics.

The Workflow Management Utility would implement or support a federated identity management system as needed to support authentication requests needed for ReCAP services.

Some of the features of the ReCAP Workflow Management Utility might include:

Business logic that determines the availability of a ReCAP item to a library user based on:

- Embedded ReCAP Customer Code of the item
- Patron institution and category

The utility would perform the processing and validation of the ReCAP request forms embedded on each of the online catalogs or discovery services. It would validate patron eligibility and calculate valid pick-up locations based on applicable parameters.

ReCAP requests should show in a patrons ILS account as would a hold or other request for library materials.

- This capability might be accomplished through a dynamic process, possibly implemented through an NCIP call that pushes a transaction to the ILS placing the item on call for the patron as part of the processing of the ReCAP request form.
- For non-local ReCAP items, this process might include the creation of a temporary item record in the ILS to support the expected subsequent circulation transaction.
- Return of items should likewise remove temporary item records

Patrons should be automatically notified when an item that they requested from ReCAP is available for pick up.
• This functionality would be accomplished by the local ILS implementation following the injection of the hold request made when the item was requested

When placing a ReCAP request a list of valid pick-up requests should be presented

• This feature would be accomplished by a dynamic call made by the ReCAP request form to the Workflow Management Module that would return the valid options for that patron category

ReCAP materials should be tracked through the entire fulfillment and return process. Patrons should receive something like a tracking ID that they can use to view the status of their request.

Materials should be scanned at each step of the fulfillment and return process, including:

• Pickup at the ReCAP facility
• Arrival at the library
• Arrival at the requested pick-up location
• Check-out to patron
• Check-in by patron
• Pick-up at library in transit to ReCAP
• Arrival at ReCAP
• Re-filed at recap into storage module

Workflow Management Utility should generate performance statistics to document percentage of transactions with on-time delivery

Workflow Management Utility should generate alerts of items that have not been delivered in expected time frames.

The Workflow Management Utility would include a Web form that patrons can use to inquire the status of a pending request. This feature would work much like the tracking systems used by parcel delivery services, displaying details of all previous steps in processing the request up to the current status.

A form would be available for library personnel to see the status of any given item or category of items.
Figure 15 General flow of fulfillment transactions

Optional Functionality or future enhancements
Each of the three institutions participating in ReCAP also participated in the Google Books mass digitization project. They also are involved in the HathiTrust, a digital repository that holds the library copies of Google Book scans, in addition to other materials. One possible element of functionality would involve a query to HathiTrust to determine if the item could be fulfilled digitally. There would be limited circumstances where the copyright status of the item would allow such digital delivery.

Technical implementation
The new technology infrastructure might be called the Enhanced ReCAP Discovery and Delivery Service. This infrastructure will not be a standalone application, but rather a set of components assembled as middleware to provide services to the strategic business application and end-user discovery services in place in the participating institutions.

The proposed functionality described will be implemented in a set of tools that will be delivered through a service-oriented architecture. While some of the functionality might involve interfaces operated by library patrons or staff, much of the environment will consist of services exposed as part of the middleware application that will be consumed in system-to-system operations. These services will exploit existing APIs or protocol responders currently or potentially available in the related integrated
library systems and discovery services. The GFA Library Archival System lacks these APIs, and part of the scope of the development will be the creation of a service layer for this application.

Although the functionality has been designed specifically for ReCAP the intent is also to make the technical environment as generalizable as possible so that it could also be adapted to other organizations. To the extent possible the services supporting the business logic will be constructed using API’s and protocols commonly used in the library and archives environment, including those defined in the Digital Library Federation sponsored ILS-DI initiative, the Mellon funded eXtensible Catalog project, NISO standards including OAI-PMH, NCIP, SIP, and SRU. APIs created specifically in support of the project should be created as RESTful Web services, authenticated and transported via SSL.

The design should be implemented in ways that provide the highest level of flexibility of possible. Many of the assumptions and business rules initially expressed to create these designs may be subject to reconsideration. To the extent possible, the business logic should be based on policy tables or parameters that can be adjusted as needed. Such flexibility will be needed for such fundamental constructs such as the definition of the ReCAP Shared Collection.

As a service-oriented application, the ReCAP Workflow Management Utility will be constructed as business logic and services that will operate on an existing middleware framework. Making use of one of the existing open source and commercial middleware environments already available would significantly reduce the effort of a development project rather than building all the low-level supporting services anew.

The selection of the low-level middleware framework to support the development of the ReCAP Workflow Management Utility requires further investigation. One possibility that stands out for consideration is Kuali Rice:

The Kuali Rice software provides an enterprise class middleware suite of integrated products that allows for applications to be built in an agile fashion. This enables developers to react to end-user business requirements in an efficient and productive manner, so that they can produce high quality business applications.

Kuali Rice is leveraged heavily by the Kuali applications but is also designed to be used in non-Kuali applications. Its services and framework pieces are designed in such a way to be applicable to multiple business domains.

Rice is built with Service Oriented Architecture (SOA) concepts in mind. Specifically, end developers are able to build robust systems with common enterprise workflow functionality, customizable and configurable user interfaces with a clean and universal look and feel, and general notification features to allow for a consolidated list of work "action items." Additionally, there are a set of services in Rice that provide identity and access management capabilities and can be used to abstract away from underlying
institution-specific identity services. All of this adds up to a re-usable development framework that encourages a simplified approach to developing true business functionality as modular applications.

[http://www.kuali.org/rice](http://www.kuali.org/rice)

The use of Kuali Rice for low-level infrastructure would provide a solid technical environment and would be consistent with any requirements for software developed under the current or follow-on grants to be made available as open source. The Kuali Foundation is closely aligned with the Andrew W. Mellon foundation, the key funding source for this ReCAP current planning effort. Kuali Rice is used by the Kuali OLE project as its low-level service bus and workflow support layer.

Other development frameworks could likewise provide the foundation of the ReCAP technical support infrastructure. IBM, for example, offers a SOA development framework called the IBM Integration Designer. [http://www-01.ibm.com/software/integration/integration-designer/](http://www-01.ibm.com/software/integration/integration-designer/)

---

**ReCAP Support System**

![ReCAP Support System Diagram](image)

*Figure 16 Simplified model of ReCAP support architecture*
Support for Electronic Document Delivery Requests

The fulfillment of electronic document delivery requests would also benefit from the enhanced technical infrastructure. While no physical tracking will be needed, tracking of the status of requests will continue to be useful. Creating transaction records would ensure that all requests are tracked and reported consistently.

The general environment as designed for request of physical materials would apply to electronic document requests, but additional requirements may be identified. To what extent can the Workflow Management Utility as conceived also interoperate with EDD systems such as Ariel?

Electronic scans made in support of ReCAP requests are currently discarded after fulfillment. Workflows could be designed that automatically transfer these scans to HathiTrust. If not already present in the HathiTrust repository, articles and book chapters could be processed through OCR and converted into formats consistent with the repository. While it seems that scans made by ReCAP staff should have similar legal status as those performed by Google, such a step would naturally require careful review. It does seem that the cumulative number of scans performed at ReCAP would represent a significant contribution to the body of digitized library content. This transfer would also need to be accomplished in such a way that does not add additional work to the ReCAP process for fulfilling EDD requests.

Detection of Duplicate Copies of Materials

The ongoing operation of the ReCAP facility will benefit by the ability to identify duplication of materials among the three institutions. Tools can be created to identify duplication among the materials currently in the facility and for new submissions. These tools can be created to be informational, allowing the partners to process duplicates according to prevailing policy or specific circumstances.

The processes designed to extract bibliographic and item information from the three ILS implementations to populate the central discovery index provide an opportunity to identify existing duplicates in the system. This process would complement the de-duplication data and analysis that has been developed through the OCLC collection analysis report and would provide an operational data set that can be used to provide automation support for any de-accessioning that may be carried out. This de-duplication routine would be a subset of the Collection Management module that would be developed as part of the ReCAP support infrastructure. The de-duplication processes will be needed to consolidate bibliographic entries and manage holding data needed for the discovery layer. The data derived from this process can be captured and further leveraged to facilitate collection management, including decision support for de-accessioning.

The programs created to extract records from each ILS and to populate the central index can extended to include algorithms to identify duplicates. Triggers for duplicates might include unique identifiers such as ISBN numbers, but would need many other match points as well. Serials would need to be processed at least at the volume level, if not by individual issues. De-duplication algorithms should be available from other projects. The population of discovery services routinely includes a de-duplication stage making it likely that program code with advanced options and configurability should not have to be created from scratch.
The other and more important consideration involves the identification of duplicate materials as they are accessioned into ReCAP. There are multiple ways in which this process can be handled:

Create a tool to be used in by library personnel as they select materials to be transferred to ReCAP. This tool might include a simple Web interface where the staff member scans in barcode or enters any other unique identifier. Once submitted, the ReCAP Duplication Tool would query the relevant targets:

- Recap Bibliographic database (middleware) (SQL query)
- Each of partner ILS, using Z 38.50 and/or NCIP
- GFA: an additional query, based on the barcode number identified for a likely duplicate could be sent to the LCA system to poll its status. This status would confirm the duplicate or it might reveal an item that has been
- WorldCat (optional, to identify other relevant libraries that hold the item)

After processing the query, the Duplicate Detection tool would return an informational page confirming that the item is unique within the ReCAP partners or listing all the duplicates that exist already in the ReCAP collection or within the partner’s active collections.

A batch version of the Duplicate Detection utility would also be created to handle large batches of materials. This material would generate a persistent electronic report, possibly held in a database table, which could be accessed by library staff to make transfer decisions based on duplication status or holdings in partner collections.

The Duplicate Processing utility would be configured according to policies in effect among the ReCAP partners. These policies would control any actions triggered by the detection of duplicates. Messaging options would include notification of both the library considering submitting the item as well as the library that submitted the existing duplicate.

Another alternative might involve extending the GFA software to perform detection of duplicates as new materials are accessioned. Building detection of duplicates into the current accessioning process presents some challenges. First, it might be counterproductive to identify materials as duplicate after they arrive at ReCAP for accessioning rather than making transfer decisions based on duplication status. It is also the case that the inventory data in the GFA software would likely be insufficient to identify duplicates without the support of the middleware to query the ILS systems to retrieve the bibliographic details.

**Next Generation Automation**

One of the key considerations for the project involves the possible upcoming transitions in the automation environments of the three partner libraries. Based on information gathered to date, it seems that none of the institutions are far enough in planning for next-generation automation systems that short-term development can target anything other than their current environment.

Both of the automation products in the mix, Millennium from Innovative Interfaces and Voyager from Ex Libris have adequate integration capabilities currently to support what is needed for the new ReCAP
technology support environment. Both have been integrated with a variety of discovery environments and have pragmatic means of interacting with discovery and fulfillment services. ILS-DI connectors have been implemented, for example, for both Millennium and Voyager. Both systems support NCIP or have other means to dynamically communicate with services related to patron and item data. Innovative offers optional APIs for Millennium to support patron and Item functionality. The eXtensible Catalog project has developed connectors for both ILS products.

The new-generation systems that the institutions are considering should offer more sophisticated interoperability mechanisms, using native APIs and Web services. Both systems follow a service-oriented architecture that should provide high levels of interoperability.

Both Sierra and Alma, however, stand at a very early stage in their development where the focus of attention has so far focused on basic functionality and not on APIs and interoperability. Princeton, even as a development partner with Alma, for example, has not been asked to test and evaluate APIs that Ex Libris has touted as a distinguishing characteristic of its new system.

In conversations with both Innovative Interfaces and Ex Libris, both organizations report a strategic commitment to delivering APIs that can be used at customer sites to achieve the interoperability and extensibility as envisioned for the ReCAP technology environment. Yet, both companies are currently focused on the basic functionality involved with early production deployments.

It seems clear that short term development will target existing ILS implementations. All work should be done in ways that could easily be adapted to other ILS products and to next-generation library services platforms such as Alma and Sierra.

The same kinds of concerns apply to the online catalog and discovery services in place. Without loading millions of ReCAP records into each ILS implementation, it will be very difficult to support broad discovery in the standard online catalogs of the three institutions. All of the enhanced request features can be accomplished in the online catalogs with discovery limited to the institution’s own ReCAP materials. It might be possible to implement a federated search component to expand discovery within these online catalogs, but even this approach is unlikely to be entirely satisfactory.

Further discussion is needed to determine if it is acceptable to offer broad discovery of the ReCAP materials through the discovery interfaces in place (BiblioCommons, Primo, and Blacklight) or if further work needs to be done to attempt to identify mechanisms other than record loading to accomplish comprehensive discovery of the Shared ReCAP collection through the online catalogs.
Future Enhancements / Feature Creep

The proposed architecture of the new generation of technology to support ReCAP should lend itself to additional capabilities. An instance of the discovery service could be implemented that addresses the entire ReCAP collection, even including the non-shareable materials in support of staff functions such as de-duplication and collection development. While the main instance of the ReCAP discovery service would consist primarily of indexing capability to integrate into existing interfaces, this staff-oriented instance would also include implementation of a search interface.

While functionality to support collection development is currently considered out of scope, the new environment is based on an architecture that could be extended to support different scopes. One scope would include all ReCAP materials; another possible scope would span the entire collections of the three institutions. There would be ways to support collaborative collection development, but the design and implementation of the technology platform should scale to the level of a union catalog of the three ReCAP institutions.

Ongoing work

This report serves as an environmental scan and a tentative exploration of functionality and technical implementation options. Vetting of the report among other ReCAP participants may expose omissions or additional functionality, workflows, or technology components that need to be evaluated. A workshop exploring the technology options was conducted on August 21, 2012. Some of the follow-up expected from the partner institutions included feedback on the functionality and general workflows proposed. Prioritization of the modules will also benefit next stages of the project that will develop cost proposals for creation of the technology infrastructure.

Additional research is needed to further refine the technology components that might form the basis of the new environment. Further work, for example, is needed to ensure the feasibility of the index integration strategy proposed, the suitability of Kuali Rice as the low-level service platform, and for Apache SOLR as the indexing platform, and whether the eXtensible Catalog toolkits provide adequate connectivity and metadata management support.

Next stages, performed in collaboration with the Technology Architect, will include the extension of the this tentative design to provide a more detailed expression of the functionality expected, the specific technical components to be used, an inventory of services to be created, and an adequate level of documentation to inform a development team that that would ultimately create the software. This next phase would also include creating estimates of costs of development and implementation and feasible timeframes which would inform the budget of the follow-on project to implement the new ReCAP technology support infrastructure.
Functional and technical requirements need to be developed regarding the work that needs to be done to create a service layer for the GFA Library Archival System. More detailed conversations need to take place with GFA technical personnel. A document needs to be created which more fully describes the work to be contracted to extend the GFA system that can serve as a Request for Proposal that would serve both as a means to determine the cost of the project and to provide the detailed requirements that would drive and assess the development. It is likely that only GFA itself would be in a position to carry out this work given that it is proprietary licensed software.
Appendix J. ReCAP High-Level Architecture 1.0
RECAP PROJECT
HIGH LEVEL ARCHITECTURE

1.0

Authors
LAKSHMI NARASIMHAN
&
VIJAYKUMAR GUNDAVARAPU
HTC Global Services, Inc.
+1 248 786 2500
lakshmi.narasimhan@htcinc.com
vijaykumar.gundavarapu@htcinc.com

Sponsor
RECAP
New York Public Library
Notice of Proprietary Information

All information contained in or disclosed by this document is confidential and proprietary to ReCAP.

By accepting this material the recipient agrees that this material and the information contained therein will be held in confidence and will not be reproduced in whole or in part without express written permission.

Distribution: Restricted

Disposal: Shred

File Name: ReCAP High Level Architecture

Copyright © 2013 ReCAP

Revision History

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Author</th>
<th>Description</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Draft</td>
<td>05/16/2013</td>
<td>Lakshmi Narasimhan &amp; Vijaykumar Gundavarapu</td>
<td>Initial Version – Draft</td>
<td>In Progress</td>
</tr>
<tr>
<td>1.0</td>
<td>5/30/2013</td>
<td>Lakshmi Narasimhan &amp; Vijaykumar Gundavarapu</td>
<td>Review comments incorporated</td>
<td>Complete</td>
</tr>
</tbody>
</table>
# Table of Contents

1. **INTRODUCTION** .................................................................................................................. 5  
   1.1 PURPOSE .......................................................................................................................... 5  
   1.2 DOCUMENT OVERVIEW ................................................................................................. 5  
   1.3 AUDIENCE ....................................................................................................................... 5  
   1.4 DEFINITION FOR ARCHITECTURE .............................................................................. 6  
   1.5 SCOPE ............................................................................................................................ 6  

2. **PROJECT GOALS** ............................................................................................................... 6  
   2.1 PROJECT SCOPE AND OBJECTIVES ............................................................................ 6  
   2.2 ARCHITECTURAL CHALLENGES WITH CURRENT SYSTEM .................................... 7  
   2.3 HOW THE NEW ARCHITECTURE ADDRESSES THE CHALLENGES .................... 7  

3. **ARCHITECTURAL OVERVIEW** ...................................................................................... 8  
   3.1 CANDIDATE ARCHITECTURE ..................................................................................... 8  
      3.1.1 Layers ...................................................................................................................... 8  
         3.1.1.1 Presentation Layer ............................................................................................ 8  
         3.1.1.2 Enterprise Services Layer ................................................................................ 8  
         3.1.1.3 Data Services Layer ......................................................................................... 9  
         3.1.1.4 Data Layer ....................................................................................................... 9  
   3.2 RATIONALE .................................................................................................................... 10  
      3.2.1 Rationale for using ReCAP Middleware database vs. ILS Transfer ....................... 10  
      3.2.2 Rationale for using Kuali RICE vs. other commercial/open source frameworks .... 10  

4. **ARCHITECTURAL VIEWS** ................................................................................................ 12  
   4.1 USE-CASE VIEW .......................................................................................................... 12  
      4.1.1 Search Shared Collection Items .............................................................................. 13  
         4.1.1.1 Architectural Significance ................................................................................ 13  
      4.1.2 Request Item ........................................................................................................... 14  
         4.1.2.1 Architectural Significance ................................................................................ 14  
      4.1.3 Validate Request .................................................................................................... 14  
         4.1.3.1 Architectural Significance ................................................................................ 14  
      4.1.4 Place Hold on Item .................................................................................................. 14  
         4.1.4.1 Architectural Significance ................................................................................ 14  
      4.1.5 Recall Item .............................................................................................................. 14  
         4.1.5.1 Architectural Significance ................................................................................ 14  
      4.1.6 Accession Item ........................................................................................................ 14  
         4.1.6.1 Architectural Significance ............................................................................... 15  
      4.1.7 DeAccession Item .................................................................................................... 15  
         4.1.7.1 Architectural Significance ............................................................................... 15  
      4.1.8 Process Borrow Direct Request .............................................................................. 15  
         4.1.8.1 Architectural Significance ............................................................................... 15  
      4.1.9 Re-file Item ............................................................................................................. 15  
         4.1.9.1 Architectural Significance ............................................................................... 15  
      4.1.10 Check Item Availability ....................................................................................... 16  
         4.1.10.1 Architectural Significance .............................................................................. 16  
      4.1.11 Get Shared Collection Records ............................................................................ 16  
         4.1.11.1 Architectural Significance .............................................................................. 16  
      4.1.12 Submit Collection Information ............................................................................. 16  
         4.1.12.1 Architectural Significance .............................................................................. 16
4.1.13 Receive Collection Updates ................................................................. 16

4.2 LOGICAL VIEW .................................................................................. 16

4.2.1 Analysis Model .............................................................................. 17

4.2.1.1 Overview ............................................................................... 17

4.2.1.2 Analysis Packages ............................................................... 17

4.2.1.3 Key Analysis Classes ................................................................ 18

4.2.1.4 Boundary Classes .................................................................. 18

4.2.1.5 Control Classes ...................................................................... 18

4.2.1.6 Entity Classes .......................................................................... 18

4.2.2 Design Model .............................................................................. 19

4.2.2.1 Architecturally Significant Design Packages .............................. 19

4.2.3 Frameworks, Patterns and Guidelines .......................................... 21

4.2.3.1 Common Patterns .................................................................. 21

4.2.3.2 Model-View-Controller (MVC) Pattern .................................... 21

4.2.3.3 Façade Pattern ....................................................................... 22

4.2.3.4 Business Delegate Pattern ...................................................... 23

4.2.3.5 Factory Pattern ....................................................................... 23

4.2.3.6 Singleton Pattern .................................................................... 24

4.2.3.7 Chain of Responsibility Pattern .............................................. 24

4.2.3.8 Value Object or Transfer Object Pattern ................................. 25

4.2.3.9 Service Locator ..................................................................... 25

4.2.3.10 Data Access Object ............................................................... 26

4.2.3.11 Inversion of Control (Spring Framework) ................................ 26

4.3 PROCESS VIEW .................................................................................. 27

4.4 DEPLOYMENT VIEW ......................................................................... 28

4.4.1 Amazon Cloud Configuration (Production Instance) .................... 29

4.4.2 Amazon Cloud Configuration (QA & Development Instance) .... 29

4.5 IMPLEMENTATION VIEW ................................................................. 30

4.5.1 Layers ....................................................................................... 30

4.5.2 Error handling .......................................................................... 31

4.6 DATA VIEW ........................................................................................ 33

5. SIZE AND PERFORMANCE .................................................................. 35

5.1 SCALABILITY .................................................................................... 35

5.2 PERFORMANCE .............................................................................. 36

6. PROPOSED DEVELOPMENT ENVIRONMENT ...................................... 37

6.1 HARDWARE .................................................................................... 37

6.2 SOFTWARE ...................................................................................... 37

7. PROTOTYPE/PROOF-OF-CONCEPTS ................................................. 39

8. QUALITY ............................................................................................. 39
1. Introduction

1.1 Purpose

Architecture represents the significant design decisions that shape a system, where significance is measured by cost of change. This document proposes high-level candidate architecture for the ReCAP middleware system. The architecture is described using a number of different views to depict architecturally-significant aspects of the system. It is intended to capture and convey the significant architectural decisions, which have been made on the system.

This version of the document presents high-level aspects of the system and the candidate architecture. As the project evolves through the development life cycle, this document will be updated to reflect the architectural decisions/design for the following (but not limited to):

- Finalized Data View
- Finalized Implementation View
- Finalized Deployment View

1.2 Document Overview

The architecture of the application is represented using the recommendations of the Rational Unified Process guidelines. This document also highlights the development environment, quality requirements and prototyping details.

The UML (Unified Modeling Language) specification of the new ReCAP middleware system has been divided into six views (Rational’s 4+1 model):

- Use Case View – illustrates and validates the architecture by presenting selected architecturally significant use cases.
- Logical View – illustrates the object model of the design. It presents an analysis model, which captures the analysis of the use cases and a design model. This view also describes the logical structure of the system and presents key structural and behavioral elements.
- Process View – illustrates the assignment of components to the operating system processes and threads.
- Implementation View - describes the physical organization of the software and its components in the production environment.
- Deployment View – illustrates the mapping of the software to the hardware and its distribution aspects.

1.3 Audience

The primary audience for this document is the Development team and QA team. The development team will use it to help aid the detailed design during the development phase and ultimately to develop the system. The QA team will use it to ensure testability and also to ensure that proper test cases are written.

Each view as presented in the “Document Overview” section primarily caters to different audience.

- Use-case view – All
- Logical view – Development team
- Process view – Software Integrators (part of the development team), QA team to understand performance and scalability bottlenecks for testing purposes
- Implementation view – Development and Deployment teams(part of development team)
• Deployment view – Deployment and Production Support teams

1.4 Definition for Architecture

The Unified Software Development Process [8] defines “Software Architecture” as

“We can think of the architecture of a system as the common vision that all the workers (i.e., developers and other stakeholders) must agree on or at least accept. The architecture gives us a clear perspective of the whole system, which is necessary to control its development.”

This document uses the architecture definitions presented by Software Architecture in Practice and the UML Modeling Language Guide: Software Architecture in Practice [7] defines “Software Architecture” as:

“The software architecture of a program or computing system is the structure or structures of the system, which comprise software components, the externally visible properties of those components, and the relationships among them.”

UML Modeling language user guide [6] defines “Software Architecture” as:

“An architecture is the set of significant decisions about the organization of a software system, the selection of the structural elements and their interfaces by which the system is composed, together with their behavior as specified in the collaborations among those elements, the composition of these structural and behavioral elements into progressively larger subsystems, and the architectural style that guides this organization---these elements and their interfaces, their collaborations, and their composition. ”

1.5 Scope

This document provides the candidate architecture for the ReCAP middleware system, which includes details for each layer (Presentation, Enterprise Services, Data Services, and Data) and the interfaces to ILS, OPAC and GFA LAS systems.

2. Project Goals

This section provides the business needs, project goals and architectural issues with the current system and explains how the new ReCAP middleware system architecture aims to address these issues. This section provides a business perspective to the architecture and establishes architectural goals, assumptions and constraints.

2.1 Project scope and objectives

“The scope of the project is to expand the vision of the ReCAP facility from a shared storage facility to a shared collection with enhanced access to the patrons of each of the participating libraries by implementing an integrated ReCAP middleware system utilizing established industry architectures. The functional requirements for ReCAP middleware will encompass all of the existing functionality, plus changes and enhancements to improve user experience and collection management.”

The ReCAP project has the following main objectives:
1. Improve visibility of ReCAP shared collection items from any participating institution in existing OPAC systems
2. Display of real-time status of items in ReCAP, including availability for request, restrictions
3. Improve services that can be embedded into the online catalog or discovery services of the participating institutions to capture and validate requests made by patrons or by library staff for ReCAP materials.
4. Provide real-time tracking for ReCAP materials requested by patrons from the time that they leave the ReCAP until they are returned for refilling.
5. Provide tools to support the management of the ReCAP collection, such as collaborative collection development, automated processing of duplicates, or designation of preservation retention.

### 2.2 Architectural Challenges with current system

1. **Shared Collection Visibility** – Items placed in shared collection by other partners is not available in the OPAC systems, limiting access of such items to patrons.
2. **Real-time Availability** – ReCAP item status is unavailable in OPAC or ILS.
3. **Real-time Request Processing** – Request processing is batch only with minimal validations and error reporting.
4. **Real-time Status Reporting** – Overall status of items between ReCAP facility and delivery locations are partially captured and distributed across disparate systems.
5. **Collection Management** – No centralized collection management is in place.

### 2.3 How the new Architecture addresses the challenges

1. **Shared Collection Visibility** – ReCAP middleware consolidates and normalizes ReCAP item and bib records from all three partners and provides nightly feeds to all partner OPAC systems. ReCAP search service provides ability to perform federated search on shared collection from OPAC.
2. **Real-time Availability** – ReCAP middleware database maintains real-time status of all ReCAP items. Item availability is provided through ReCAP middleware API.
3. **Real-time Request Processing** – ReCAP middleware maintains validation rules and item status. Request submitted through OPAC forms are validated real-time, processed and recorded in ReCAP middleware database. Users receive confirmation or validation error messages in real-time enabling them to resubmit a valid request.
4. **Real-time Status Reporting** – ReCAP middleware consolidates a complete view of item status across GFA and ILS systems into middleware database. Consolidated status can be leveraged for tracking and analytics.
5. **Collection Management** – ReCAP middleware implements centralized automated collection classification algorithm. Middleware provides user interfaces for manual workflow steps such as withdrawal of preservation copies.
3. **Architectural Overview**

3.1 **Candidate Architecture**

![Candidate Architecture Diagram]

**Figure 1- Candidate Architecture**

### 3.1.1 Layers

The architecture includes four distinct layers:
- Presentation Layer
- Enterprise Services Layer
- Data Services Layer
- Data Layer

#### 3.1.1.1 Presentation Layer

The presentation layer deals with user interface aspects of the system. Presentation layer will leverage Kuali Rapid Application Development (KRAD), a framework providing reusable solutions and templates. KRAD is built upon industry standard jQuery libraries providing out-of-box UI components, validations and accessibility to RICE middleware.

#### 3.1.1.2 Enterprise Services Layer

The enterprise services layer encapsulates specific business rules, which are made available to the presentation layer. The presentation layer requests enterprise services, which are then fulfilled by this
layer. The architecture envisages providing a seamless enterprise service layer communicating with internal data stores and 3rd party services. The data access layer supports the enterprise service layer by serving the data required.

Enterprise Services is based on a Service Oriented Architecture and leverages Kuali Service Bus (KS
B), for service integration. Services will be designed as java spring-based services and will be published on the service bus as remote asynchronous calls. Transaction services will be published as SOAP services and lookup services will be published as RESTFul Services.

Features such as Service Discovery, Quality of Service, Security, Monitoring and Messaging are available as out-of-box features in Kuali Service Bus and can be leveraged during implementation as required.

3.1.1.3 Data Services Layer

The data services layer provides fundamental services to fulfill the business needs (fulfilled through enterprise services) such as Search, Request Item, etc. The data services layer serves data required by enterprise services. Data services support both relational database and Solr.

Services implementing data access to relational database will leverage Java Persistence Architecture (JPA), providing separation of object persistence and data access logic from a particular persistence mechanism (relational database) in data layer. This approach provides the flexibility to change the applications persistence mechanism without the need to re-engineer application logic that interacts with the data layer. Persistence classes are developed following the object-oriented idiom including association, inheritance, polymorphism, composition, and collections. This framework provides the flexibility to express queries in its own portable SQL extension, as well as in native SQL, or with object-oriented criteria.

Services implementing data access to Solr / Lucene search will wrap the Solr RESTFul API's to provide features such as search, filter, sort and navigation.

3.1.1.4 Data Layer

The data layer serves as the data store for all persistent information in the system including the relational database and search engine indexes.

RDBMS data layer will comprise of MySQL cluster. RDBMS data layer will be accessed only from the data access layer via Data Access Objects (DAOs). RDBMS cluster architecture allows a single physical database to be accessed by concurrent instances running across several different CPUs. The proposed data layer will be composed of a group of independent servers or nodes that operate as a single system. These nodes have a single view of the distributed cache memory for the entire database system providing applications access to more horsepower when needed while allowing computing resources to be used for other applications when database resources are not as heavily required. In the event of a sudden increase in traffic, proposed system can distribute the load over many nodes, a feature referred to as load balancing. In addition to this, proposed system can protect against failures caused by unexpected hardware, operating system or server crashes, as well as processing loss caused by planned maintenance. When a node failure occurs, connection attempts can fail over to other nodes in the cluster, which assumes the work of the failed node. When connection failover occurs and a service connection is redirected to another node, users can continue to access the service, unaware that it is now provided from a different node.

A single Solr instance can support more than one index using Solr cores (single index per core). A single large index can be a performance overhead. SolrCloud distributes a single index on different machines,
commonly referred as shards. All shards of the same index making one large index are referred as collection. While collection supports index scaling, it does not provide redundancy. Replication of shards provides redundancy and fault tolerance.

Zookeeper maintains the SolrCloud, by distributing the index across shards and federating the search through the collection. SolrCloud uses leaders and an overseer. In the event of leader or the cluster overseer failure, automatic fail over will choose new leaders or a new overseer transparently to the user and they will seamlessly takeover their respective jobs. Any Solr instance can be promoted to one of these roles.

3.2 Rationale

3.2.1 Rationale for using ReCAP Middleware database vs. ILS Transfer

Cross loading MARC records representing holdings of other ReCAP institutions into each ILS was considered as a design alternative. This option would have introduced significant costs and support burdens for partners. In some cases current ILS systems capacity or license thresholds would be exceeded. All three partners are planning to replace their current ILS systems in near future. This option would have incremented the data migration effort from existing ILS to the new ILS.

Loading ReCAP bibliographic and item records in middleware database provides a centralized repository for shared collection without impacting the ILS systems at partner institutions. Middleware database is needed to store the entire item and bib records of all there partner’s private and shared collections. Also it is needed to record all the incoming requests from patrons and to maintain the transactions. Hence .middleware database is required irrespective of the decision to synchronize bibliographic and item records. This option comes at a marginal increase to the implementation and ongoing maintenance costs.

Hence the approach of loading shared bibliographic and item records to ReCAP middleware database is recommended over cross loading MARC records to partner ILS.

3.2.2 Rationale for using Kuali RICE vs. other commercial/open source frameworks

Major components required to support the architecture includes Service Bus, Rules Engine, Workflow Engine, Authentication and Authorization and User Interface/Experience framework.

While several open source projects such as JBoss, Spring, and JQuery presented compelling components, these components have to be integrated by the project team to provide a seamless platform for ReCAP middleware.

Kuali RICE framework presents the benefits of open source such as no license costs and vendor dependencies. The framework leverages several industry standard frameworks such as Spring, JQuery, etc. and provides an enterprise grade end-to-end integrated framework well suited for ReCAP middleware development.

Rice is built on a Service Oriented Architecture (SOA) providing common enterprise workflow functionality, customizable and configurable user interfaces with a clean and universal look and feel, and general notification features to allow for a consolidated list of work “action items.” Additionally, there are a set of services in Rice that provide identity and access management capabilities and can be used to abstract away from underlying institution-specific identity services. All of this adds up to a re-usable development framework that encourages a simplified approach to developing true business functionality as modular applications.
Kuali Service Bus (KSB) provides service management and routing functionalities. Workflow and Messaging domain is taken care by Kuali Enterprise Workflow (KEW) and Kuali Enterprise Notification (KEN). Kuali Identity Management (KIM) provides services for authentication and authorization management. Also it has Kuali Rules Management (KRM) for business rule development and execution as well as information delivery and analysis.

Kuali foundations commitment to provide and support enterprise scale framework for the higher education and academic library community makes Kuali RICE a compelling choice for this project.
4. **Architectural Views**

4.1 **Use-Case View**

The Architecturally significant Use Cases identified during the High Level Architecture definition are listed below. The diagram provides the model for architecturally significant use cases.

![Use-Case Diagram](image)

*Figure 2 – Architecturally Significant Use-Case View*
The following table lists the actors (user or system) interacting with the system.

<table>
<thead>
<tr>
<th>No.</th>
<th>Actor</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Patron</td>
<td>A library patron is someone who uses a library, a university student or a city resident. Typically, this person gets a library card, browses the available books, CDs, DVDs, etc.</td>
</tr>
<tr>
<td>2</td>
<td>Library Staff</td>
<td>A library employee, who is responsible for a collection of specialized or technical information about items and management of items in a library.</td>
</tr>
<tr>
<td>3</td>
<td>ReCAP Staff</td>
<td>A Person responsible for day to day activities at GFA facility including accessioning, deaccessioning, filing, re-filing, etc.</td>
</tr>
<tr>
<td>4</td>
<td>OPAC</td>
<td>An Online Public Access Catalog (often abbreviated as OPAC or simply Library Catalog) is an online database of materials held by a library. Example: Bibliocommons, CLIO.</td>
</tr>
<tr>
<td>5</td>
<td>ILS</td>
<td>An integrated Library System (ILS) is an enterprise resource planning system for a library, used to track items owned, orders made, bills paid, and patrons who have borrowed. Example: Millennium, Voyager</td>
</tr>
</tbody>
</table>

The following lists the architecturally significant Use Cases.

<table>
<thead>
<tr>
<th>No.</th>
<th>Use Case Name</th>
<th>Architecture Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Search Shared Collection Items</td>
<td>Complex</td>
</tr>
<tr>
<td>2</td>
<td>Request Item</td>
<td>Complex</td>
</tr>
<tr>
<td>3</td>
<td>Validate Request</td>
<td>Simple</td>
</tr>
<tr>
<td>4</td>
<td>Place Hold on Item</td>
<td>Complex</td>
</tr>
<tr>
<td>5</td>
<td>Recall Item</td>
<td>Complex</td>
</tr>
<tr>
<td>6</td>
<td>Accession Item</td>
<td>Medium</td>
</tr>
<tr>
<td>7</td>
<td>Deaccession Item</td>
<td>Medium</td>
</tr>
<tr>
<td>8</td>
<td>Process Borrow Direct Request</td>
<td>Complex</td>
</tr>
<tr>
<td>9</td>
<td>Re-file Item</td>
<td>Complex</td>
</tr>
<tr>
<td>10</td>
<td>Check Item Availability</td>
<td>Complex</td>
</tr>
<tr>
<td>11</td>
<td>Get Shared Collection Records</td>
<td>Medium</td>
</tr>
<tr>
<td>12</td>
<td>Submit Collection Information</td>
<td>Medium</td>
</tr>
<tr>
<td>13</td>
<td>Receive Collection Updates</td>
<td>Medium</td>
</tr>
</tbody>
</table>

A brief description of the architecturally significant use cases has been listed below. Each of the use case description includes key business rules and includes reasons for architectural significance.

### 4.1.1 Search Shared Collection Items

In this use-case the patron will search the OPAC for institution items as well as shared collection items placed by other ReCAP partners. Search for an item in OPAC will initiate search to OPAC’s index and ReCAP index. The two search results will be merged to include shared collection items in the search results.

#### 4.1.1.1 Architectural Significance

- Core Functionality
- Complexities –Includes collecting bibliographic and item data from all three partners,
normalizing and indexing the data and providing offline feed or API for OPAC systems.

4.1.2 **Request Item**
In this use-case a patron will request a ReCAP item by submitting the request through a web form. The form will submit the request to middleware API, which will invoke other use-cases to process the request.

4.1.2.1 **Architectural Significance**
- Core Functionality
- Complexities – ReCAP middleware will interact with GFA LAS, ILS and OPAC to process the request. It will create a temporary item record in one of the three applicable ILS depending upon the patron.

4.1.3 **Validate Request**
Request item use-case will invoke this use-case to validate the request for requested item, delivery location, delivery type, etc. Upon successful validation control will be returned to the main use-case with a confirmation message and upon unsuccessful validation an error message will be returned.

4.1.3.1 **Architectural Significance**
- Core Functionality
- Complexities – ReCAP middleware will validate against ReCAP circulation policies and item availability in the middleware database.

4.1.4 **Place Hold on Item**
In this use-case a patron will place hold against an item whose status is currently unavailable.

4.1.4.1 **Architectural Significance**
- Core Functionality
- Complexities – ReCAP middleware will maintain a single hold queue for all the partner institutions in a first-in, first-out basis. The hold queue will be automatically propagated to all applicable ILS systems.

4.1.5 **Recall Item**
In this use-case a patron/library staff will recall an item whose status is currently unavailable.

4.1.5.1 **Architectural Significance**
- Core Functionality
- Complexities – ReCAP middleware will interact with owning or borrowing institution ILS to send the Recall request and maintain the queue in middleware database.

4.1.6 **Accession Item**
In this use-case Library Staff will upload bib and item records for new ReCAP items. ReCAP middleware will interface with GFA LAS to check the accessioned item status and then apply accessioning algorithm. Applying accessioning algorithm will result in one of the following three scenarios. A valid collection code will be assigned after which the item will be a shared collection item. It might also result in duplicates in
which case, the items might be placed under institutional access instead of shared access. Detecting duplicates might also result in libraries withdrawing their items. So the proposed accessions are the items which are sent by Library staff for accessioning and actual accessions are the items which are assigned collection code after running the accessioning algorithm. Item information with assigned collection codes will be returned back to the owning libraries through SFTP drop. ReCAP staff also participates in the accessioning of an item.

4.1.6.1 Architectural Significance
- Core Functionality
- Complexities – Accessioning algorithm will be run every time an item is accessioned in ReCAP. Accessioning algorithm includes a tie-breaker to cover most of the scenarios. Match and normalize disparate bib and item data across three partner ILS and GFA LAS. The item barcodes and applied circulation codes data will be returned to owning partner ILS.

4.1.7 DeAccession Item
In this use case Library Staff will initiate a request to deaccession an item through staff interfaces. Based on the collection code a manual approval workflow will be triggered to deaccession the item. ReCAP staff also participates in the deaccessioning of an item.

4.1.7.1 Architectural Significance
- Core Functionality
- Complexities – ReCAP middleware will run accessioning algorithm to reassign circulation codes for other items after deaccessioning an item. A review/approval workflow will be implemented to manage preservation collections.

4.1.8 Process Borrow Direct Request
This use-case will be invoked by ReCAP staff to process a Borrow direct request. The staff will scan the barcode in the Borrow direct request or enter one if barcode not available. Upon matching the barcode the staff can invoke the request item use case by clicking the confirmation button.

4.1.8.1 Architectural Significance
- Core Functionality
- Complexities – ReCAP middleware will provide a thick client interface for barcode scanning to the ReCAP staff. The solution will maintain existing workflow for ReCAP staff and integrate the solution to middleware.

4.1.9 Re-file Item
In this use-case middleware will poll GFA LAS for re-filed items periodically, if an item is re-filed and has no hold or recall queue against it, its status will be changed to available. If a hold/recall queue exists the item will be processed for the first patron in the queue.

4.1.9.1 Architectural Significance
- Core Functionality
- Complexities – ReCAP middleware will actively poll GFA LAS to get the current status of the item. Once the item is checked-in (GFA), ReCAP middleware will process the item for next patron in queue and update corresponding ILS system.
4.1.10 Check Item Availability

In this use-case OPAC will request for a real-time availability status of an item from ReCAP middleware. Middleware API will return the status from the index which is maintained in sync with the transaction database.

4.1.10.1 Architectural Significance
- Core Functionality
- Complexities – Real-time Item status will be provided through search API which is maintained in sync with the ReCAP database. Update search engine index without performance degradation.

4.1.11 Get Shared Collection Records

In this use-case OPAC systems will retrieve the other partner’s shared collection records from SFTP server. The bib and item record will be normalized during inbound process and will be de-normalized during the outbound process to fit each partner’s needs. The outbound records will be limited to other institution’s shared collection items.

4.1.11.1 Architectural Significance
- Core Functionality
- Complexities – De-normalizing feeds for five OPAC systems.

4.1.12 Submit Collection Information

In this use-case partner ILS system will provide collection information, new accessioned and updates to bibliographic data through SFTP upload. Middleware will process data from all partners, normalize the data and ingest into middleware database. The normalized data will be updated to ReCAP index.

4.1.12.1 Architectural Significance
- Core Functionality
- Complexities – Normalizing bib and item data from three ILS systems and de-normalizing feeds for five OPAC systems

4.1.13 Receive Collection Updates

In this use-case the ILS systems will retrieve collection updates from ReCAP middleware through SFTP drops. ReCAP middleware will de-normalize the data sets and provide updated collection information of item records pertinent to requesting institution only.

4.1.13.1 Architectural Significance
- Core Functionality
- Complexities – Identifying the collection update and provide offline export of owning library items only.

4.2 Logical View

The Logical View consists of two models: Analysis model and Design Model.
4.2.1 Analysis Model

4.2.1.1 Overview

The Analysis Model provides a view of the requirements from the system’s perspective. The requirements are refined, structured and the resulting elements are organized into logical groups of similar functionality called analysis packages.

The analysis model contains View-of-participating-classes (VOPC) which map out control, entity and boundary classes.

4.2.1.2 Analysis Packages

The Analysis Model contains the following main packages. The following diagram shows the overall package structure and dependencies:

<table>
<thead>
<tr>
<th>Package Name</th>
<th>Package Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Aggregation</td>
<td>This package contains classes which are responsible for consolidating and normalizing bib and item data from all three partner feeds.</td>
</tr>
<tr>
<td>Data Distribution</td>
<td>This package contains classes responsible for de-normalizing data from ReCAP middleware database and then distributes shared collection data to all 3 partners through SFTP uploads.</td>
</tr>
<tr>
<td>Search &amp; Discover</td>
<td>This package contains classes which handle all the search requests from OPAC systems.</td>
</tr>
<tr>
<td>Validate Request</td>
<td>This package contains classes which are responsible for validating any incoming request.</td>
</tr>
<tr>
<td>Request Item</td>
<td>This package contains classes responsible for processing a ReCAP request.</td>
</tr>
<tr>
<td>Hold Item</td>
<td>This package contains classes responsible for processing and maintaining hold queue.</td>
</tr>
</tbody>
</table>
Recall Item | This package contains classes responsible for processing a recall request.
---|---
Accession Item | This package contains classes responsible for accessioning new items.
DeAccession Item | This package contains classes responsible for Deaccessioning existing items.
Reports | This package contains classes responsible for generating reports for Library staff.

### 4.2.1.3 Key Analysis Classes

Significant analysis classes are described below.

### 4.2.1.4 Boundary Classes

Boundary classes represent the interface between the system and the actors.

<table>
<thead>
<tr>
<th>Class Name</th>
<th>Class Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFA Client</td>
<td>A class responsible for invoking SOAP service to interact with GFA LAS for item status.</td>
</tr>
<tr>
<td>NCIP Client</td>
<td>A class responsible for invoking NCIP responders to update/create item details.</td>
</tr>
<tr>
<td>Search Service</td>
<td>A class publishing Services related to bib and item records and its real time status.</td>
</tr>
<tr>
<td>Barcode Client</td>
<td>A class responsible for handling events related to barcode scanning (Borrow Direct requests)</td>
</tr>
</tbody>
</table>

### 4.2.1.5 Control Classes

Control classes represent classes that co-ordinate flow between the entities and the boundary classes.

<table>
<thead>
<tr>
<th>Class Name</th>
<th>Class Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentication Controller</td>
<td>Controller Class which handles all the incoming requests for authentication.</td>
</tr>
<tr>
<td>Search Controller</td>
<td>Controller Class which handles all the search requests for shared collection items.</td>
</tr>
<tr>
<td>Request Controller</td>
<td>Controller Class which handles all the transactions for item requests.</td>
</tr>
<tr>
<td>Workflow Controller</td>
<td>Controller Class which handles all the workflow requests such as Accessioning and Deaccessioning items.</td>
</tr>
<tr>
<td>Report Controller</td>
<td>Controller class which handles all the report requests.</td>
</tr>
</tbody>
</table>

### 4.2.1.6 Entity Classes

Entity classes represent information and associated behavior that must be stored. They are usually persistent. Important entity classes are listed below:
### Class Name | Class Description
--- | ---
Bib | A bibliographic record is an entry being a uniform representation and description of a specific content item in a bibliographic database (or a library catalog), containing data elements required for its identification and retrieval, as well as additional supporting information, presented in a formalized bibliographic format
Item | An item record represents a physical piece in the library.
Request | Entity that contains all the details for every single request received by ReCAP middleware

#### 4.2.2 Design Model

##### 4.2.2.1 Architecturally Significant Design Packages

The application has been partitioned into four layers:

- Presentation Layer
- Enterprise Services Layer
- Data Services Layer
- Data Layer

The presentation layer deals with presentation aspects of the system. The enterprise service layer isolates business rules and the data service layer from the presentation layer. The enterprise service layer implements common services such as “Search”, “Retrieve” etc. The Data Service Layer deals with data.
The presentation layer has been further sub-divided into "System" and "Report" packages. The System package contains the system administration specific implementation of the interfaces. The report package contains the report specific components.

**Figure 4 - Package Hierarchy**
Enterprise Services Layer:
The enterprise services layer contains the “Service” components. The layer uses the “delegate” pattern to delegate to the appropriate architecture (RDBMS or Solr). The underlying services are accessed via a “Facade” bean. Any bean that acts as a facade would contain a delegate class within its package.
The following are sample list of packages that exist within this layer:
Search & Discover – This package contains classes which handle all the search requests from OPAC systems.
Request Item – This package contains classes handling all the logic to process a ReCAP request.
Hold Item – This package contains classes responsible for processing and maintaining hold queue
Recall Item – This package contains classes responsible for processing a recall request.
Accession Item – This package contains classes responsible for accessioning new items
DeAccession Item – This package contains classes responsible for Deaccessioning existing items.
Data service layer contains RDBMS and Search Packages. RDBMS package consists of all classes which interact with the relational database and Search Packages consists of classes which interact with SOLR.

4.2.3 Frameworks, Patterns and Guidelines
Application frameworks are a promising technology for reifying proven software designs and implementations in order to reduce the cost and improve the quality of software.
A framework is a reusable, "semi-complete" application that can be specialized to produce custom applications. In contrast to earlier OO reuse techniques based on class libraries, frameworks are targeted for particular business units (such as data processing) and application domains (such as user interfaces)

4.2.3.1 Common Patterns
All the diagrams presented in this section are copyright of their respective creators and ReCAP project will use most it not all the below patterns during its implementation.

4.2.3.2 Model-View-Controller (MVC) Pattern
The proposed ReCAP architecture will use MVC model for its core framework. The diagram below (courtesy Sun Microsystems) explains the concepts behind MVC:
• **Model**: The model represents enterprise data and the business rules that govern access to and updates of this data. Often the model serves as a software approximation to a real-world process, so simple real-world modeling techniques apply when defining the model.

• **View**: The view renders the contents of a model. It accesses enterprise data through the model and specifies how that data should be presented. It is the view's responsibility to maintain consistency in its presentation when the model changes. This can be achieved by using a push model, where the view registers itself with the model for change notifications, or a pull model, where the view is responsible for calling the model when it needs to retrieve the most current data.

• **Controller**: The controller translates interactions with the view into actions to be performed by the model. In a stand-alone GUI client, user interactions could be button clicks or menu selections, whereas in a Web application, they appear as GET and POST HTTP requests. The actions performed by the model include activating business processes or changing the state of the model. Based on the user interactions and the outcome of the model actions, the controller responds by selecting an appropriate view.

#### 4.2.3.3 Façade Pattern

The façade pattern is used in the design at many points. Most significant use is via the Service Implementations, where a Service Implementation Class (business service implementation) acts as a façade to the business layer (hiding the business layer complexities) and also provides a simpler interface for the clients to work with.

The diagram below highlights the details of Façade Pattern:
The intent of this pattern is to hide complex underlying structural details with a simpler interface providing following benefits.

- Provides a simpler interface for the clients to work with
- Reduces number of objects that the client needs to work with
- Promotes weak coupling

### 4.2.3.4 Business Delegate Pattern

The Business Delegate pattern can be used to reduce coupling between presentation-tier clients and business services. The Business Delegate hides the underlying implementation details of the business service, such as lookup and access details of the business tier components. The lookup service could be implemented using the “Service Locator” pattern.

It is normally implemented by defining a business interface, which is implemented by a delegate class and the business component (if the component is being created afresh. If not, then the delegate acts as a façade).

The diagram below illustrates this pattern:

This model will be used wherever multiple implementations are possible for an interface.

### 4.2.3.5 Factory Pattern

Factory Method is a pattern used for object creation. This pattern helps model an interface for creating an object, which at creation time can let its sub-classes, decide the class to be instantiated. This is called a Factory Pattern since it is responsible for "Manufacturing" an Object. The Factory Pattern promotes loose coupling by eliminating the need to bind application-specific classes into the code.
The application layer framework will use this pattern to manage the ingestion factory.

4.2.3.6 Singleton Pattern

Singleton pattern is used to ensure a class has only one instance, and provide a global point of access to it. It also encapsulates "just-in-time initialization" or "initialization on first use". The application framework will use this pattern to initialize single instances of all configuration properties.

4.2.3.7 Chain of Responsibility Pattern

Chain of Responsibility pattern is used to avoid coupling the sender of a request to its receiver by giving more than one object a chance to handle the request. This is achieved by chaining the receiving objects and passing the request along the chain until an object handles it. Each object in the chain launches and leaves requests with a single processing pipeline that contains many possible handlers thus creating an object-oriented linked list with recursive traversal.

The application framework will use this pattern to manage several independent steps of output creation.
4.2.3.8 Value Object or Transfer Object Pattern

In typical applications there is a need to get several properties or exchange business data between tiers. This exchange requires multiple round trips over a network and could result in poor performance. To improve the performance the business data could be encapsulated into a “Value” or “Transfer” object and passed to the service. Similarly the service could return a “Value” object, rather than having the client make several “get” calls. Typically value objects have a VO suffix.

The application framework will use this pattern to exchange data across tiers. The following diagram illustrates this pattern:

![Value Object or Transfer Object Pattern Diagram](image)

4.2.3.9 Service Locator

Service locator pattern is used to abstract the complexities of initializing all services. Multiple clients can reuse the Service Locator object to reduce code complexity, provide a single point of control, and improve performance by providing a caching facility. This pattern reduces the client complexity that results from the client's dependency on and need to perform lookup and creation processes, which are resource-intensive. To eliminate these problems, this pattern provides a mechanism to abstract all dependencies and network details into the Service Locator.

![Service Locator Diagram](image)
4.2.3.10 Data Access Object

Data Access Object (DAO) pattern is used to abstract and encapsulate all access to the data source. The DAO manages the connection with the data source to obtain and store data.

The DAO implements the access mechanism required to work with the data source. The data source could be a persistent store like an RDBMS, an external service like a B2B exchange, a repository like an LDAP database, or a business service accessed via CORBA Internet Inter-ORB Protocol (IIOP) or low-level sockets. The business component that relies on the DAO uses the simpler interface exposed by the DAO for its clients. The DAO completely hides the data source implementation details from its clients. Because the interface exposed by the DAO to clients does not change when the underlying data source implementation changes, this pattern allows the DAO to adapt to different storage schemes without affecting its clients or business components. Essentially, the DAO acts as an adapter between the component and the data source.

The application framework will use this pattern to encapsulate all data store calls from services.

4.2.3.11 Inversion of Control (Spring Framework)

Inversion of Control or IoC is one of the techniques used to wire services or components to an application program. By definition, IoC is "A software design pattern and set of associated programming techniques in which the flow of control of a system is inverted in comparison to the traditional interaction mode." Simply stated, in IoC, instead of an application calling the framework, it is the framework that calls the components specified by the application.

However, IoC is a broad and generic term. The aspect of IoC that the Spring Framework uses is "Injection of required resources or dependency at Run-time into the dependent resource," which is also known as Dependency Injection. Hence, the service provided by the IoC container of Spring is Dependency Injection.
4.3 Process View

The diagram below shows the process views from within various layers with respect to the Web container and clients interacting with the system:

![Diagram showing process views with layers and interactions](image-url)

**Figure 5 - Overall Process View**
4.4 Deployment View

The deployment view of the ReCAP system shows the physical nodes on which the system executes and the assignment of the system processes to the nodes. The system can be deployed on different hardware configurations.

![Deployment View Diagram](image)

**Figure 6 - Deployment View**

Please refer to “Size and Performance” section for more details.
### 4.4.1 Amazon Cloud Configuration (Production Instance)

<table>
<thead>
<tr>
<th>Component</th>
<th>Name</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database</td>
<td>RDS - Heavy Utilization Extra Large Single AZ (15 GB of memory, 8 ECU (4 virtual cores with 2 ECU each), 64-bit platform, High I/O Capacity, Provisioned IOPS Optimized: 1000Mbps)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>DB Single AZ Storage (250 GB storage)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Provisioned IOPS</td>
<td>1</td>
</tr>
<tr>
<td>Web Server</td>
<td>EC2 - Heavy Utilization High CPU Extra Large (7 GiB of memory, 20 EC2 Compute Units (8 virtual cores with 2.5 EC2 Compute Units each), 1690 GB of local instance storage, 64-bit platform)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>EBS Optimization Fee for the Extra Large Instance</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>EC2 - Heavy Utilization Standard Medium (3.75 GiB of memory, 2 EC2 Compute Units (1 virtual core with 2 EC2 Compute Units each), 410 GB of local instance storage, 32-bit or 64-bit platform)</td>
<td>1</td>
</tr>
<tr>
<td>Storage</td>
<td>EBS Storage (250 GB storage)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Provisioned IOPS</td>
<td>1</td>
</tr>
<tr>
<td>Backup Storage</td>
<td>S3 Snapshot of EBS Volumes (500 GB storage)</td>
<td>1</td>
</tr>
<tr>
<td>Load Balancer</td>
<td>Elastic Load Balancer</td>
<td>1</td>
</tr>
</tbody>
</table>

### 4.4.2 Amazon Cloud Configuration (QA & Development Instance)

<table>
<thead>
<tr>
<th>Component</th>
<th>Name</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database</td>
<td>RDS - Heavy Utilization Large Single AZ (7.5 GB memory, 4 ECUs (2 virtual cores with 2 ECUs each), 64-bit platform, High I/O Capacity, Provisioned IOPS Optimized: 500Mbps)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>DB Single AZ Storage (250 GB storage)</td>
<td>1</td>
</tr>
<tr>
<td>Web Server</td>
<td>EC2 - Heavy Utilization Standard Large (7.5 GiB of memory, 4 EC2 Compute Units (2 virtual cores with 2 EC2 Compute Units each), 850 GB of local instance storage, 64-bit platform)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>EBS Optimization Fee for the Extra Large Instance</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>EC2 - Heavy Utilization Standard Medium (3.75 GiB of memory, 2 EC2 Compute Units (1 virtual core with 2 EC2 Compute Units each), 410 GB of local instance storage, 32-bit or 64-bit platform)</td>
<td>1</td>
</tr>
<tr>
<td>Storage</td>
<td>EBS Storage (250 GB storage)</td>
<td>1</td>
</tr>
<tr>
<td>Backup Storage</td>
<td>S3 Snapshot of EBS Volumes (500 GB storage)</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: EC2 Compute Unit (ECU) – One EC2 Compute Unit (ECU) provides the equivalent CPU capacity of a 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor.
4.5 Implementation View

The implementation view shown here is only a starting point and will be refined in during development phase.

4.5.1 Layers

The layers, packages and its hierarchy are represented using the following diagrams. The package hierarchy starts with standard Java namespace compliant structure and then is divided into two sub-packages for the different layers viz. src – Application (Services, Data Access) and Web Content – Presentation.
4.5.2 **Error handling**

Error handling will be implemented by leveraging the Exceptions feature of the Java language. The following guidelines are suggested when dealing with exceptions.

<table>
<thead>
<tr>
<th>No.</th>
<th>Exception Guideline</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Exceptions should not create additional package dependencies</td>
<td>Assume that a client class in package A accesses a class in package B. The class in package B should not throw an exception that belongs to package C (which is used by package B). This produces dependencies between package A and C. (This rule may not apply where Package C is a standard and stable package, such as java.io.)</td>
</tr>
</tbody>
</table>
### Exceptions by package.

If a package's classes throw any exceptions, the package should have its own top-level checked exception. The package should then define exception subclasses for any exceptions that may be handled differently by clients. Good models for this paradigm can be found in the Java packages `java.io`, `java.sql` and `javax.naming`. Sometimes a package is clearly a "sub-package" of another package. In such a case, the sub-package's exceptions can extend the parent package's exceptions. In example of such a sub-package is `java.nio.charset` (whose exceptions extend `java.io.IOException`).

### No blind catches of Exception

A class is responsible for knowing what exceptions it may encounter, and it must treat each exception individually. If the handling of many exceptions is identical, it could be extracted into helper methods.

### No empty catch-blocks

At the very least, a catch-block should contain an assertion that it should never be reached or a comment stating that it is irrelevant.

### Write sensible throws clauses

Fewer (<3) the number of exceptions thrown, better it is. Always throw exceptions that make sense to the calling class, if not wrap that exception in another, which more closely captures the error type.

### Chaining Exceptions

Always chain exceptions so that the root cause of the error is available for logging it into the error/system log file. This is very useful for diagnosing errors in production environment.

### Use Message Catalogs for easy localization

Use message catalogs for message text of an exception, whose message is directly presented to the end user. This will help the application to be localized or internationalized by just adding another message catalog.

Applications when encountering an exception should always log it to the Application/System log. Lower level components should avoid writing to an error/system log.
4.6 Data View

The following table lists all the entities along with their definition:

<table>
<thead>
<tr>
<th>No.</th>
<th>Entity</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Item</td>
<td>An item record represents a physical piece in the library</td>
</tr>
</tbody>
</table>

Figure 8 – Entity Relationship
<table>
<thead>
<tr>
<th></th>
<th>Bib</th>
<th>A bibliographic record is an entry being a uniform representation and description of a specific content item in a bibliographic database (or a library catalog), containing data elements required for its identification and retrieval, as well as additional supporting information, presented in a formalized bibliographic format</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Location</td>
<td>Entity that contains list of valid delivery locations</td>
</tr>
<tr>
<td>4</td>
<td>Partner</td>
<td>List of names and contacts of the ReCAP partners</td>
</tr>
<tr>
<td>7</td>
<td>Request</td>
<td>Entity that contains all the details for every single request received by ReCAP middleware.</td>
</tr>
<tr>
<td>8</td>
<td>Status</td>
<td>Entity that contains details of all the statuses applicable for items in ReCAP middleware</td>
</tr>
<tr>
<td>9</td>
<td>RequestStatus</td>
<td>Entity that contains details and timelines of all statuses associated to a ReCAP request.</td>
</tr>
<tr>
<td>10</td>
<td>RequestHistory</td>
<td>Archive Entity for completed Request (without any patron information).</td>
</tr>
<tr>
<td>11</td>
<td>RequestStatusHistory</td>
<td>Archive table for RequestStatus entity.</td>
</tr>
<tr>
<td>12</td>
<td>CirculationCode</td>
<td>Entity that contains details about various types of circulation policies for shared collection items</td>
</tr>
<tr>
<td>13</td>
<td>Restriction</td>
<td>Entity that contains details about policy restrictions for shared collection items</td>
</tr>
</tbody>
</table>
5. Size and Performance

5.1 Scalability

This section explains how the architecture aims to achieve scalability.

![Figure 9 - Scalability]

The figure above presents how the architecture aims to be scalable. There are several elements that contribute to scalability:

- **Network or HTTP load balancers** – These appliances or devices would perform load balancing of HTTP and other protocol specific servers. The actual mechanism of load balancing will depend upon specific device and could include mechanisms like round robin, cookie sniffing etc. In a web environment, the load balancers will balance the load between web servers. The web server maintains state (user specific). This typically means that once a session is established, a user is redirected to the same web server. Relatively inexpensive servers (nodes) could be used for the web servers. Redundant servers could be used to provide high-availability.

- **Web Server Cluster** – The Web server cluster appears to the client application (Browser or Service Client) as a single server. The Web server provides clustering capability. Although session state could be replicated, it could result in performance hits. The architecture presents design using sticky sessions to provide high availability and fault tolerance without compromising performance. The Web server cluster scales by adding more nodes to the cluster. The applications will need no change when the cluster scales. Again, relatively inexpensive servers could be used to enable linear scaling.

- **RDBMS cluster** – The RDBMS servers will be clustered to provide scalability. The RDBMS cluster appears as a single server to the user of the database. The RDBMS product takes care of data replication and clustering challenges.

- **Solr Cloud** – Solr Cloud creates a cluster of Solr servers representing two different shards of a collection (complete index). While shards provide distributive scaling, shard replication provides fault tolerance. Zookeeper takes care of data replication and clustering challenges.
5.2 Performance

This section presents how the architecture addresses performance related issues.

This section presents various design patterns used to achieve performance.

Basic performance metrics are latency and throughput. Latency is measured as the time elapsed between request and response and throughput as the number of requests handler per second. In an ideal world, the latency should not increase and throughput should scale linearly as the load increases.

Performance related issues needs to be investigated at various points of the architecture. Some common elements that should be subjected to performance tuning are:

- Middleware – Middleware technologies like ESB or other distributed technologies are primary candidates for performance tuning as issues like network round-trips and network latency could become critical.
- Database – Database access and processes like joins and sorts are candidates for performance tuning.
- Search Engine – Solr caching is an candidate to improve search performance by leveraging cached queries and results
6. Proposed Development Environment

6.1 Hardware

The core technology will be Java and an implementation of a servlet container. MySQL will be the relational data store and apache SOLR will be the search engine. Application development environment will be hosted in Amazon cloud on Linux. Each developer will have his or her own development setup on PC or Mac and access source code that is stored in a common source control repository such as SVN or CVS.

It is assumed that Partners (NYPL, Princeton and Columbia) will be hosting the development, QA and production environments of NCIP Responders and GFA LAS.

6.2 Software

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Eclipse Juno 4.2</td>
<td>Eclipse is a multi-language software development environment comprising a base workspace (Eclipse Public License (EPL))</td>
</tr>
<tr>
<td>2</td>
<td>MySQL 5.6</td>
<td>Open source relational database management system (RDBMS) that runs as a server providing multi-user access to a number of databases (GNU General Public License)</td>
</tr>
<tr>
<td>3</td>
<td>Tomcat 7.0</td>
<td>Apache Tomcat is an open source web server which provides pure Java HTTP web server environment for Java code to run. (Apache License)</td>
</tr>
<tr>
<td>4</td>
<td>Kuali RICE 2.2.3</td>
<td>Kuali RICE comprises of a suite of middleware programs (workflow, messaging, identity management), interfaces and Web services around a service bus (Educational Community License)</td>
</tr>
<tr>
<td>5</td>
<td>Apache SOLR 4.2.1</td>
<td>SOLR is an open source enterprise search platform written in Java and runs as a standalone full-text search server within a servlet container. (Apache License)</td>
</tr>
<tr>
<td>6</td>
<td>Apache Quartz 2.1.7</td>
<td>Quartz is a full-featured, open source job scheduling service that can be integrated with, or used alongside virtually any Java application (Apache License)</td>
</tr>
<tr>
<td>7</td>
<td>ProFTPD 1.3.5rc2</td>
<td>ProFTPD is an secured FTP server exposing a large configuration options to the user (GNU General Public License)</td>
</tr>
<tr>
<td>8</td>
<td>Jenkins 1.511</td>
<td>Jenkins is a server-based system running in a servlet container providing open source continuous integration features. (Massachusetts Institute of Technology (MIT) License)</td>
</tr>
<tr>
<td>9</td>
<td>JUnit 4</td>
<td>JUnit is a unit testing framework for the Java Programming language (Common Public License)</td>
</tr>
<tr>
<td>10</td>
<td>SOAP UI 4.5.1</td>
<td>SoapUI is an open source web service testing application for service-oriented architectures (SOA) and provides functionalities like web service inspection, invoking, development, simulation and mocking, functional testing, load and compliance testing (GNU Lesser General Public License (LGPL))</td>
</tr>
<tr>
<td>11</td>
<td>Maven 3.0.5</td>
<td>Maven is a build automation tool used primarily for Java projects (Apache License)</td>
</tr>
<tr>
<td>12</td>
<td>Java Platforms (Java 7)</td>
<td>Java is a set of several computer software products and specifications that together provide a system for developing application software and deploying it in a cross-platform</td>
</tr>
<tr>
<td></td>
<td></td>
<td>computing environment (Freeware)</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>13</td>
<td>Selenium IDE (1.10.0)</td>
<td>Selenium is a portable software testing framework which provides record/playback tool for authoring tests. (Apache License)</td>
</tr>
<tr>
<td>14</td>
<td>Motorola Scanner SDK</td>
<td>A framework providing a single programming interface across multiple programming languages and across multiple system environments for all Motorola scanners</td>
</tr>
<tr>
<td>15</td>
<td>Java Swing</td>
<td>Swing is the primary Java GUI widget toolkit developed to provide a more sophisticated set of GUI components</td>
</tr>
</tbody>
</table>
7. **Prototype/Proof-of-Concepts**
   The proposed architecture is based on proven industry standard components, hence no specific prototype or proof-of-concept developed. Core technology will be Java and an implementation of a servlet container.

8. **Quality**
   Extensibility: The proposed architecture is based on Components adhering to well-defined interfaces and industry based standards (J2EE), so adding new features and implementing new components will not require extensive rework of existing components.

   Reliability: The architecture uses standard J2EE architecture and will have the capabilities of Load balancing. Fail-over mechanisms will improve the reliability of the system.

   Portability: The Enterprise Service Layer abstracts the platform and protocol specific implementation of common services into a generic set of interfaces. Clients use these interfaces to access the services are not tied to protocol/platform. It will be possible to port applications to different protocols/platforms by implementing the Generic set of interfaces on the client protocol/platform.

   By complying with J2EE standards, it will be possible to port the application to multiple server vendors.
### Glossary

<table>
<thead>
<tr>
<th>No.</th>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>ReCAP</td>
<td>Research Collections and Preservation Consortium (ReCAP) is a storage facility for all the shared collection items.</td>
</tr>
<tr>
<td>2.</td>
<td>UML</td>
<td>Unified Modeling Language (UML) is a standardized, general-purpose modeling language in the field of software engineering.</td>
</tr>
<tr>
<td>3.</td>
<td>GFA LAS</td>
<td>Generation Fifth Applications - Library Archive System is an inventory management system from Generation Fifth Applications which catalogs and controls archival storage of shared collection items in the ReCAP facility.</td>
</tr>
<tr>
<td>4.</td>
<td>Patrons</td>
<td>Users of the System who place a request for a shared collection item/items.</td>
</tr>
<tr>
<td>5.</td>
<td>ReCAP Middleware</td>
<td>The central component of this architecture which handles search, discover, request processing, reporting and collection management for a shared collection item by interacting with other components in the system.</td>
</tr>
<tr>
<td>6.</td>
<td>Shared collection items</td>
<td>Any item stored in the ReCAP facility which can be requested by any of the partners.</td>
</tr>
<tr>
<td>7.</td>
<td>Institution items</td>
<td>Any item stored in the ReCAP facility which can be requested only by the owning partner.</td>
</tr>
<tr>
<td>8.</td>
<td>Refiling</td>
<td>Refiling is a process of re-shelving the item in the ReCAP facility.</td>
</tr>
<tr>
<td>9.</td>
<td>Item Records</td>
<td>An item record represents a physical piece in the library</td>
</tr>
<tr>
<td>10.</td>
<td>Bib Records</td>
<td>A bibliographic record is an entry being a uniform representation and description of a specific content item in a bibliographic database (or a library catalog), containing data elements required for its identification and retrieval, as well as additional supporting information, presented in a formalized bibliographic format</td>
</tr>
<tr>
<td>11.</td>
<td>Federated Search</td>
<td>Federated search is an information retrieval technology that allows the simultaneous search of multiple searchable resources</td>
</tr>
<tr>
<td>12.</td>
<td>SOA</td>
<td>Service-oriented architecture (SOA) is a flexible set of design principles used during the phases of systems development and integration</td>
</tr>
<tr>
<td>13.</td>
<td>Temporary item record</td>
<td>An item record which is created in any of the partners ILS which is temporary in nature and requires deletion in the near future</td>
</tr>
<tr>
<td>14.</td>
<td>Circulation policies</td>
<td>Rules related to circulation of any given item to a patron</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>15.</td>
<td><strong>Hold</strong></td>
<td>To place a hold on an item means to reserve it. An item that is checked out may have a hold placed on it by another patron who wishes to use it. When the item is returned, the library will contact the patron who is waiting so they may come in and check it out.</td>
</tr>
<tr>
<td>16.</td>
<td><strong>Recall</strong></td>
<td>Recall is a special type of request by a library to a borrower for the return of a borrowed item before the due date.</td>
</tr>
<tr>
<td>17.</td>
<td><strong>Owning institution</strong></td>
<td>Any institution which borrows an item which belongs to itself is called an owning institution.</td>
</tr>
<tr>
<td>18.</td>
<td><strong>Borrowing institution</strong></td>
<td>Any institution which borrows an item which belongs to other partners is called an borrowing institution.</td>
</tr>
<tr>
<td>19.</td>
<td><strong>Accession</strong></td>
<td>The process of recording an item and its location in the ReCAP facility into the GFA LAS system.</td>
</tr>
<tr>
<td>20.</td>
<td><strong>Barcodes</strong></td>
<td>A barcode is an optical machine-readable representation of data relating to the object to which it is attached. Every item such as books or films and location such as aisle, shelf, and bin has a unique barcode.</td>
</tr>
<tr>
<td>21.</td>
<td><strong>Circulation code</strong></td>
<td>A unique identifier for every item which dictates its circulation policy.</td>
</tr>
<tr>
<td>22.</td>
<td><strong>Collection Code</strong></td>
<td>A code which is assigned to determine the scope of the sharing of an item.</td>
</tr>
<tr>
<td>23.</td>
<td><strong>Customer Code</strong></td>
<td>A unique identifier which is currently used in GFA to identify a group to which an item belongs.</td>
</tr>
<tr>
<td>24.</td>
<td><strong>Staff interfaces</strong></td>
<td>User Interface screens for Library staff</td>
</tr>
<tr>
<td>25.</td>
<td><strong>Borrow Direct</strong></td>
<td>Borrow Direct is an interlibrary borrowing service offered by all of the Ivy League Universities except Harvard.</td>
</tr>
<tr>
<td>26.</td>
<td><strong>Source Control Repositories</strong></td>
<td>It’s a space set aside to maintain code base for the ReCAP middleware project.</td>
</tr>
<tr>
<td>27.</td>
<td><strong>SVN(Apache Subversion)</strong></td>
<td>It’s a type of Source control repository.</td>
</tr>
<tr>
<td>28.</td>
<td><strong>CVS(Concurrent Version System)</strong></td>
<td>It’s a type of Source control repository.</td>
</tr>
<tr>
<td>No.</td>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>-----</td>
<td>--------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>29.</td>
<td>NCIP Responders</td>
<td>National Information Standards Organization Circulation Interchange Protocol (NCIP) is a protocol that is limited to the exchange of messages between and among computer-based applications to enable them to perform functions necessary to lend and borrow items, to provide controlled access to electronic resources, and to facilitate cooperative management of these functions. It's a mechanism by which ReCAP middleware communicates with any of the ILS which belongs to the participating institutions.</td>
</tr>
<tr>
<td>30.</td>
<td>Fault tolerance</td>
<td>Fault-tolerant describes a computer system or component designed so that, in the event that a component fails, a backup component or procedure can immediately take its place with no loss of service.</td>
</tr>
<tr>
<td>31.</td>
<td>Cluster</td>
<td>A cluster is a group of servers and other resources that act like a single system and enable high availability and, in some cases, load balancing and parallel processing.</td>
</tr>
<tr>
<td>32.</td>
<td>Shard</td>
<td>A shard is a term which is used in SolrCloud a feature in Apache SOLR which enables high fault tolerance of SolrCores.</td>
</tr>
<tr>
<td>33.</td>
<td>Load Balancer</td>
<td>Load Balancer achieves high fault tolerance by distributing incoming requests across one or more web servers.</td>
</tr>
<tr>
<td>34.</td>
<td>Deaccession</td>
<td>It is a process of withdrawing an item from ReCAP facility</td>
</tr>
<tr>
<td>35.</td>
<td>RDS</td>
<td>Amazon Relational Database Service (Amazon RDS) is a web service that makes it easy to set up, operate, and scale a relational database in the cloud</td>
</tr>
<tr>
<td>36.</td>
<td>AZ</td>
<td>Multi-AZ deployment is for enhanced data durability and availability</td>
</tr>
<tr>
<td>37.</td>
<td>IOPS</td>
<td>The ability which Amazon provides to specify or provision the I/O capacity needs is called IOPS.</td>
</tr>
<tr>
<td>38.</td>
<td>EC2</td>
<td>Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud</td>
</tr>
<tr>
<td>39.</td>
<td>EBS</td>
<td>Amazon Elastic Block Store (EBS) provides block level storage volumes for use with Amazon EC2 instances</td>
</tr>
<tr>
<td>40.</td>
<td>Elastic Load Balancing</td>
<td>Elastic Load Balancing automatically distributes incoming application traffic across multiple Amazon EC2 instances</td>
</tr>
<tr>
<td>41.</td>
<td>S3</td>
<td>Amazon S3 is storage for the Internet. Amazon S3 provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the web</td>
</tr>
</tbody>
</table>
Appendix B  References

<table>
<thead>
<tr>
<th>No.</th>
<th>Reference</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>ReCAP Grant Overview.doc</td>
<td>ReCAP Team</td>
</tr>
<tr>
<td>3.</td>
<td>ReCAP Technology Workshop.ppt</td>
<td>Marshall Breeding</td>
</tr>
<tr>
<td>4.</td>
<td>ReCAP Shared Collection Policies.doc</td>
<td>Lizanne Payne</td>
</tr>
<tr>
<td>5.</td>
<td>ReCAP Workflows.doc</td>
<td>HTC Team</td>
</tr>
<tr>
<td>6.</td>
<td>ReCAP NCIP.doc</td>
<td>HTC Team</td>
</tr>
<tr>
<td>7.</td>
<td>ReCAP Library Archival System API Requirements.doc</td>
<td>HTC Team</td>
</tr>
</tbody>
</table>
Appendix C  Naming and Coding Standards

Project will adopt the following industry standard naming and coding standards:

Code Conventions for the Java TM Programming Language

Guidelines, Patterns, and code for end-to-end Java applications

J2EE Patterns Catalog